RECOGNIZED IN "Y" CATEGORY BY

ISSN Print: 3005-8007
ISSN Online: 3005-8015
Volume 3
Issue 2
July - December 2025

UCP

Journal of Engineering & Information Technology

University of Central Punjab

ISSN:
3005-8015 (Online)
3005-8007 (Print)
Vol. 3, Issue 2
(July - December 2025)

(UCP-JEIT)
UCP Journal of Engineering & Information Technology
HEC Recognized (Y- Category)

Volume 3
Issue 2

Faculty of Information Technology & Computer Sciences
&
Faculty of Engineering

University of Central Punjab, Lahore, Pakistan

Advisory Board

International Members

Dr. Muhammad Saadi
(Nottingham Trent University, UK)

Dr. Muhammad Ramzan
(Saudi Electronic University, KSA)

Dr. Nasir Rajpoot
(University of Warwick, UK)

Dr. Ali Nasir
(King Fahad University of Petroleum and Minerals, KSA)

Dr. Adnan Qureshi
(Birmingham Newman University, UK)

Dr. Touseef Tahir
(University of Roehampton, London, United Kingdom, UK)

Dr. Safdar Ali
(Jeju National University, South Korea)

Dr. Ahmad Usman
(University of New South Wales, Australia)

Dr. Neel Kanwal
(Oslo University Hospital, Norway)

National Members

Subscription Charges National: PKR 1000 per issue
International: US$ 200 per issue

Acknowledgment

The Editorial Board of the UCP Journal of Engineering and Information Technology extends heartfelt appreciation to all those who have played crucial roles in bringing Volume 3, Issue 2 to fruition. We sincerely recognize the invaluable contributions of our esteemed researchers/authors, whose dedication to advancing knowledge has enriched this inaugural edition.

We also extend our gratitude to the diligent reviewers whose expertise and insightful feedback have ensured the quality and rigor of the articles published herein. Your commitment to the peer-review process is deeply valued.

Furthermore, we thank all individuals involved in the publication process, including editorial staff, copyeditors, and designers, whose unwavering support and tireless efforts have been indispensable.

Without the collective dedication of these individuals, the publication of Volume 3, Issue 2 of the UCP Journal of Engineering and Information Technology would not have been possible. We anticipate continued collaboration and the exploration of new frontiers in the realm of engineering and information technology.

Warm regards,
Dr. Muhammad Amjad Iqbal
Editor-in-Chief
UCP Journal of Engineering and Information Technology

Disclaimer

The views expressed in these articles are solely those of the respective authors and do not necessarily reflect the views of the Editorial Board or the management and staff of the University of Central Punjab. While every effort has been made to ensure the accuracy of the information provided by the authors, the Editorial Board does not accept any responsibility for any errors or omissions or breach of copyrights, if any.

Every effort has been made to ensure the accuracy and reliability of the information presented in the articles. However, the Editorial Board and the University of Central Punjab make no representations or warranties regarding the completeness, accuracy, or suitability of the content. Readers are encouraged to exercise their judgment and discretion when interpreting and applying the information contained in these articles.

The UCP Journal of Engineering & Technology is committed to upholding the highest standards of academic integrity and ethical publishing practices. Any concerns, questions, or requests for clarification related to the content published in this journal should be directed to the respective authors, who bear full responsibility for their work.

We appreciate your understanding of this disclaimer and hope that you find the content within this journal informative and thought-provoking.

UCP JOURNAL OF ENGINEERING AND TECHNOLOGY

Table of Contents

Date available online: 09-01-2026
Vol. 3, Issue 2 (July – December 2025)

DOI: https://doi.org/10.24312/ucp-jeit.03.02.459

Control Strategies for Smart Charging of Electric Vehicles from a Grid Perspective: A Review

Muhammad Anique Aslam[1]
[1]Department of Electrical Engineering, University of Engineering and Technology, Lahore-54890, Pakistan (e-mail: maniqueaslam@uet.edu.pk)
Corresponding author: Muhammad Anique Aslam (e-mail: maniqueaslam@uet.edu.pk)

ABSTRACT
Electric vehicles have emerged as an alternative way to reduce fossil fuel consumption, which is the cause of increasing environmental, economical and geopolitical problems. This paper reviews the strategies for charging electric vehicles smartly from the viewpoint of the grid. These strategies are classified into three categories. The strategies at the component level discuss the necessary aspects of batteries, their charging methods, and chargers for smart charging purposes. The strategies on the system level are discussed under the heads of unidirectional and bidirectional power flow strategies. Unidirectional power flow strategies manage the power flow from the grid to electric vehicles for their charging. The bidirectional power flow strategies, apart from charging the electric vehicles, also use their battery storage for grid support. Also, the strategies that can be deployed at the operational level are discussed. These strategies, on the one hand, tend to alleviate the stressful impacts of increasing the load of charging the electric vehicles on the grid, and on the other hand, use the energy storage capability of the electric vehicles for grid support.

INDEX TERMS Electric Vehicles, Smart Charging, Power Flow Control of Electric Vehicles, Centralized Control, Decentralized Control

I. INTRODUCTION

Fossil fuels have been the main source of energy throughout the growth of human civilization. Increasing industrialization, technological advancements, and machine dependent lifestyle over the past few decades have stressed fossil fuels to a dangerous level. This has resulted in various environmental, economical and geopolitical problems. The greenhouse gas emissions have increased to a hazardous level. The prices of fossil fuels are increasing and becoming more and more shaky. Above all, the demand for fossil fuels, especially oil, has resulted in terrible peace-related problems, leading to the usage of oil as an economic weapon and the instability of oil-producing countries. So, naturally, the trend has shifted towards the use of alternative sources to meet human needs [1], [2]. The usage of Electric Vehicles (EVs) is one of the attractive options for this purpose.

The usage of EVs reduces oil consumption, resulting in less greenhouse gas emissions. Also, the noise pollution is reduced. It reduces oil imports of a country, resulting in an improved economy. The cost per kilometer for an electric drive is less than that of an internal combustion engine. So, energy is used more efficiently. The energy stored in the batteries of EVs can be used to support the grid in terms of voltage and frequency regulation, peak load shaving, and tracking of Renewable Energy Sources (RESs). As a result, the number of EVs is increasing continuously [3]–[5].

From the grid perspective, EVs act as a load while charging. Studies have shown that the environmental, economical and grid-related benefits of EVs can be achieved if they are charged smartly with respect to the grid. If not charged smartly, a fleet of EVs may increase the peak load. This results in increased power demand, higher transmission losses, heating of transmission equipment, and ultimately high costs [6], [7]. With deregulated electricity markets, EVs should be charged smartly. Otherwise, they are of no economic benefit to the owner [8]–[10]. An EV powered by a coal-based power plant produces more pollution than an ordinary fossil fuel-based vehicle [11]. In short, EVs would do more harm than good if not charged smartly [12].

This paper reviews smart charging strategies of EVs from the grid perspective. The aim is to reduce the burden of adding an extra load of vehicle charging to the grid, as well as to use the storage capacity of the battery for grid support. The strategies are described under three major categories.

In Section II, component-based strategies are discussed. In Section III, strategies at the system level are described. In Section IV, strategies at the operational level are described. Finally, in Section V, the conclusions of the whole discussion are drawn, and an outlook is presented.

II. STRATEGIES AT THE COMPONENT LEVEL

No matter how smart the charging strategies are, nothing can be gained if the EVs are not able to cope with these strategies. Therefore, the components of the EVs should be able to comply with the smart charging strategies. This section discusses different aspects of the batteries, their

charging methods, and chargers that are essential for smart charging.

From the grid perspective, batteries should have high efficiency, high energy density, high charging and discharging power, and smooth charging and discharging characteristics. High efficiency reduces energy losses. High energy density imparts flexibility of storage. High charging and discharging power make it possible to charge the battery rapidly during off-peak hours and deliver large amounts of power to the grid when required. Smooth charging and discharging characteristics are desirable for maintaining good power quality [13].

Initially, lead-acid batteries were used, but these were dropped due to low energy density and environmental hazards. Then came the Nickel batteries with higher energy densities as compared to the lead-acid batteries. But these batteries have low efficiency, high self-discharge, and memory effect. Nowadays, lithium-ion batteries are used. These batteries have relatively high energy and power density and are capable of fast charging. Research is going on to improve the batteries from the grid and customer perspective [14], [15].

There are different ways to charge the battery. The most common method is constant voltage charging. In this method, the voltage is kept constant during the charging. The current is very high at the start and gradually falls to a very small value. The problem with this method is that it requires very high power at the start. The constant current method maintains a constant current during charging by changing the charging voltage. This method requires a complex method of monitoring temperature, voltage, and time to determine the cut-off. A better choice is the constant current constant voltage method. In this method, initially the battery is charged at constant current (battery voltage rises), and when the voltage reaches a predefined value, the charging method is shifted to constant voltage (now the current falls). This method is used for fast charging [16].

Instead of providing continuous voltage or current, these may be provided in the form of pulses. The width of the pulse is adjusted to meet the charging rate. A certain rest period is provided between the pulses to allow the chemical reaction to keep pace with the charging, thus avoiding the gas formation. This effect is strengthened by providing negative pulses. The selection of an appropriate charging method depends on local conditions like battery characteristics, charging circuits, driving routine, and grid constraints [14].

Charging is done through specialized power electronic circuits called chargers, which may be built inside the vehicle (on board) or outside (off board). On-board chargers are small, of low power rating, and used for slow charging. Off-board chargers are bigger in size, of high-power rating, and usually used for fast charging. These chargers use different control techniques to implement different charging processes and special circuits to lessen the grid impact of vehicle electrification. Typical examples include filters to reduce harmonics and snubbers to reduce inductive voltage spikes. The choice of a charger depends on the battery charging characteristics, driving schedule, and grid constraints [17].

III. STRATEGIES AT SYSTEM LEVEL

This section discusses the strategies that can be opted for on the system level for charging the EVs smartly from the grid frame of reference. Such strategies can be categorized into unidirectional and bidirectional power flow strategies as described below.

A. UNIDIRECTIONAL POWER FLOW STRATEGIES

These strategies treat EVs as loads taking electricity from the grid and charging the EVs. They are broadly classified into centralized and decentralized strategies [11], [18]. Some examples are as follows.

1) Centralized Strategies

A central unit controls the charging of each EV. Centralized (also known as direct) strategies are simple to implement but involve high computational effort, extensive communications, and large delays. Also, there are issues of data privacy and hacking. So, these strategies are not appropriate for large systems [11], [18]. Some of the commonly used strategies are discussed below.

A Simple Strategy for a Charging Station

In a simple charging strategy, a centralized communication system inputs some data each time a new EV arrives, such as the arrival and departure times of the EV, the state of charge (SoC) of each battery, the capacity of the battery, and the extent to which the battery should be charged. This data is used to formulate an optimization problem to minimize the power losses under the constraint of charging the battery to the desired SoC within the given time schedule without exceeding the maximum power limit of the charging station. In this way, optimized charging schedules and charging rates are determined. Such a non-linear optimization problem can be solved by sequential quadratic optimization [19], [20].

Fuzzy Logic-Based Strategy

The fuzzy logic technique uses linguistic variables to define a system, which are the words of a natural language, e.g., the linguistic variable for an air conditioning system may be defined as “temperature". Each linguistic variable is decomposed into various terms, e.g., cold, warm, etc., to qualify it. These variables are then quantified using membership functions, e.g., a numerical value is assigned to “cold" temperature. This process is called fuzzification. The interaction of these variables is assessed through different rules by an inference engine, e.g., if the temperature is warm, a command for cooling should be issued. Defuzzification of these assessments determines the output [21], [22].

Fuzzy logic based charging controller can be used to ensure a minimum network voltage while charging the EVs.The required input linguistic variables are the minimum bus voltage (obtained by power flow solution), SoC of the batteries (provided by the communication system between the EV and the battery), and electricity price (provided by the utility). These inputs are fuzzified

and assessed through knowledge-based rules by the inference engine to provide fuzzy charging levels. Defuzzification of these fuzzy charging levels results in crisp charging levels of the batteries. If these charging levels are maintained, the network voltage does not fall below a minimum value (usually 0.9 p.u.). For example, charging levels are reduced at peak load when the system is more vulnerable to voltage drop [23], [24].

As this algorithm is based on linguistic variables and general rules of system behaviour, it can be easily extended. As an example, the Vehicle to Grid (V2G) option may be added by introducing a “discharge" linguistic variable, which can be used to control the discharge of batteries for the grid support if surplus storage is available [24], [25].

Valley Filling Algorithm

The off-peak hours appear as a valley in the load profile of a network. Stress on the grid caused by the charging of EVs can be reduced by charging the EVs during the off-peak hours. Such a strategy is known as the valley-filling algorithm, which can be carried out in the following steps [26]–[28].

1) In the first step, the total charging power required by the EVs at each time step is estimated. This can be done by developing some stochastic models based on historically available data. Then the surplus power at the k^{th} time step (P_{surp}^{k}) is calculated as

$$P_{surp}^{k} = P_{conv}^{max} - P_{conv}^{k} \qquad (1)$$

where (P_{conv}^{max}) is the maximum conventional load and (P_{conv}^{k}) is the conventional load at the kth time step. After that, the capacity margin index at the kth time step (CM^{k}) is calculated as

$$CM^{k} = \frac{P_{surp}^{k}}{P_{demand}^{k}} \qquad (2)$$

Where (P_{demand}^{k}) is the charging power demanded by the EVs at the k^{th} time step and is equal to the sum of the charging powers of all the EVs connected at that time step. The time slot with the highest capacity margin is selected to charge the EVs. This ensures that the deepest point of the so-called load valley is filled first.

2) The charging priority index at the k^{th} time step for the n^{th} EV (CP^{k}) is calculated as

$$CP_{n}^{k} = \begin{cases} \frac{E_{n}^{k}}{(T_{n}^{k} \times \Delta t) P_{n}}, \text{if } I_{n}^{s} \le k \le I_{n}^{e} \\ 0, \text{else} \end{cases} \qquad (3)$$

where E_{n}^{k} is the remaining charging energy required at the k^{th} time step for the n^{th} EV, T^{k} is the remaining number of time intervals at the k^{th} time step for the n^{th} EV, Δt is the duration of one time slot, and P_n is the power of the charger of the n^{th} EV. Moreover, I_{n}^{s} and I_{n}^{e} denote the serial number of the time step of the connection and disconnection of the EV, respectively. The EV with a higher charging priority index means it has a high priority for charging in a given time slot, and vice versa. It can be seen that the EVs that are more discharged and/or have less charging time are given high priority. If the surplus power is enough to charge all the EVs in the selected time slot, all the EVs are connected. Otherwise, EVs are connected according to their charging priority.

3) The charging energy required and the time left for each EV are determined. If all the vehicles have zero charging energy required and/or the end of the time is reached, the program is terminated. Otherwise, the next iteration begins with the first step.

It should be noted that the calculations of (P_{surp}^{k}) use (P_{conv}^{max}). This ensures that the valleys are filled no higher than the peak value of the conventional load. The underlying assumption is that the EVs can be charged by using the energy available in the gaps between (P_{conv}^{max}) and (P_{conv}^{k}) But if some vehicles remain uncharged at the end of the cycle, a value higher than (P_{conv}^{max}) The value should be used. The lower this value, the lower the stress on the grid. One way to optimize this value is the dichotomy method as described in [29].

B. DECENTRALIZED STRATEGIES

In decentralized (also known as indirect, local, or distributed) strategies, each part of the system, particularly EVs, takes part in decision-making. So, computations and communications are reduced as compared to the centralized strategies. This makes these strategies attractive for large fleets of EVs [11], [18]. Some of the strategies are discussed below

Offline Heuristic or Rule-Based Strategy

The algorithm of such a control strategy determines the hours with the lowest electricity price and the charging power patterns to charge the battery in that particular time span without exceeding the load limit of the house. Specific case studies for price and peak load reduction by using this algorithm can be found in [20] and [30]. This algorithm is mostly used for simple systems and does not take into account the charging of all the vehicles in a particular network [20]. It has a high computational time, especially for complex systems [24]. It is a decentralized control and does not take into account the charging of all the vehicles in a particular network [20], [31].

A typical offline heuristic algorithm takes into account the daily load profile of a house, total power allowed by the utility, energy prices, and the arrival and departure hours of EVs. Analytical relations are used for the calculations of the battery parameters, e.g., SoC, voltage, current, etc.

First of all, the time duration for which the EV is available for charging is determined by the arrival and departure times. This time duration is sampled into time slots of equal length. The power available for charging the EV is calculated considering the power allowed by the utility and the losses of the charger. Different charging powers can be set for the EVs. Then the time

slots are sorted in ascending order of the energy prices. The time slot with the lowest energy price is selected for charging the EV. Then the current SoC of the EV is determined using analytical expressions. If the current SoC exceeds the desired SoC, the algorithm terminates. Otherwise, the voltage and the current are determined from the analytical expressions of the battery. If the battery current exceeds the nominal current, the battery is charged at the nominal current. Otherwise, the battery charges at the calculated current. Afterwards, the SoC is calculated, and the algorithm starts at the next time slot with the next lowest price. In this way, the charging is done at the lowest priced time slots. So, the charging price is minimized, and the peak load is avoided to the maximum extent [30].

A Price-Based Routing Mechanism for Charging Stations

Charging patterns of EVs are randomly distributed in temporal and spatial domains. This puts a non-uniform stress on charging stations. For example, a charging station at a particular site may be more loaded at a particular time than the other one. This leads to inefficient service of charging stations, high power losses, and congestion situations from the grid point of view, as well as inconvenience for the customers [32].

To avoid all these, a routing strategy can be employed. When the vehicle arrival rate at a particular charging station exceeds a specified limit, an increased price is offered by the charging station. This will encourage the customers to go to a nearby station, thus increasing the uniformity of load distribution. For each diverted vehicle, a penalty is imposed on the charging station as well. This is done to ensure the best efforts of the charging station to satisfy the customers. With this vehicle diversion, a communication system is designed to communicate between the vehicles and charging stations about the available locations and prices. A game theoretic model is developed where the operator of charging stations acts as one player (leader) and EVs act as another set of players, which respond to the former player (followers). Each player opts for certain actions (called "'strategies" in game theory) which result in certain outcomes (called "payoffs" in game theory). The strategy of the leader, i.e., operator of the charging stations, is to offer prices to earn maximum profit (leader payoff) by maximizing the number of customers and minimizing the diversions, keeping in view the grid constraints. In response to the leader, the followers, i.e., EVs, opt for a strategy of picking those charging stations where charging is least expensive (follower payoff) [33], [34].

Multi-Agent System-Based Strategy

A multi-agent system can be used for charging a large number of EVs (in the range of millions) in a decentralized manner. This strategy considers the EV charging system as a set of autonomous agents. An agent is an entity (physical or virtual) that senses its environment and reacts in a predefined manner to attain certain goals. In a multi-agent system, various agents interact with one another following certain rules to achieve specialized goals. A properly designed multi-agent system is robust (i.e., tolerant to faults) and modular (i.e., new agents can be added for enhanced abilities) [35], [36].

In a typical implementation, the system can be classified into three agents, namely charging stations, responsive EVs, and unresponsive EVs. Responsive EVs are those that can adjust their charging schedules in accordance with external constraints, e.g., energy prices, voltage limitations, etc. Unresponsive EVs have rigid charging schedules. The algorithm is carried out in the following steps [11].

1) In the 1st step, the arrival of a new EV is monitored. If there is a new EV, its charging is planned by referring to the 3rd step. If it is the first time step of the algorithm cycle (usually one day), the forecasting is done by executing the 2nd step.

2) In the 2nd step, the forecasting of renewable energy generation and the demand of unresponsive EVs is made. This can be based on previously available data. The conventional load (i.e. without EVs) profile comes from the distribution grid operator. Then the total power demand on conventional resources at each time step for each feeder is given by:

$$\begin{aligned} &Scheduled\ Responsive\ EV\ Load \\ &+\ Forecasted\ Unresponsive\ EV\ Load \\ &+\ Forecasted\ Conventional\ Load \\ &-\ Forecasted\ Renewable\ Energy\ Generation \end{aligned}$$

and the virtual energy price for each time step for each feeder is given by:

$$\frac{Power\ Demand}{Power\ Rating\ of\ the\ Feeder} \times Profit\ Factor$$

Profit factor can be linear, quadratic, or any other function, depending on revenue targets. It should be noted that this price is a virtual price and does not reflect the actual utility price. It can be seen that the virtual price increases with the demanded power. Such a pricing strategy encourages the EVs to charge at low price time steps, which are the time steps of off-peak loads and/or high renewable energy generation.

3) The 3rd step decides the charging schedule of each responsive EV on first come first serve basis. The objective is to minimize the product of the instantaneous charging power demand of the EV and virtual cost at that time step over the specified duration of charging.

The constraint is that the sum of the instantaneous charging power demand of the EV in the specified duration should be equal to the desired charging capacity i.e. the particular EV should be charged to the desired capacity in the available duration. Moreover, the instantaneous charging power demand of the EV should not exceed the nominal power rating of the charging station.

After the EV is scheduled, the power demand and energy price for each time step and each feeder are calculated again, as done in the second step. If

such an update of energy prices is not done, each new incoming EV will prefer to get charged at the lowest energy price points. If this is allowed to go on, the load at these points will continue to increase, and hence the stress caused by these points on the grid will increase as well. In the worst case, these valley points may become the peak load points. Moreover, such sequential updates would incentivize the early-coming responsive vehicles.

4) After scheduling each responsive EV, the network is continuously monitored. This can be done by having measurements in real time or performing a power flow analysis. If all the measurements, e.g., voltages, thermal limits, etc., are within the specified limits, the monitoring is continued until the end of the algorithm cycle is reached, and the algorithm starts again from the first step.

Meanwhile, if some new EV comes to the grid, its schedule is determined as stated above. In case something wrong happens resulting in unacceptable variations of the voltages or thermal limits, etc., the previously determined power demand and charging schedules are nullified. This may be the result of some unexpected change in production or demand. The remaining charging power demands of all the vehicles are determined. The charging station determines the power required to be rescheduled to solve the problem. Each responsive EV is rescheduled again. After rescheduling a vehicle, network conditions are monitored. If the problem is solved, no further EV is rescheduled. Otherwise, rescheduling of the next EV is done. This continues until the rescheduled power is zero or there is no EV left. As such, a condition is not the fault of the customer, no extra charges are applied for rescheduling. Some algorithms calculate the schedules at each time step to avoid such network problems, but this gives a high computational load to the algorithm.

C. BIDIRECTIONAL POWER FLOW STRATEGIES

Due to the presence of batteries, EVs act as spatially and temporally distributed energy storage. The idea of bidirectional power flow strategies is to use this available storage from the grid perspective, along with charging the EVs. The power of batteries can be used for maintaining the frequency and voltage, i.e., regulating the active and reactive power flow, preventing the line losses and transformer stress by providing local generation, providing the spinning reserve, harmonic filtering, tracking the RESs, and peak load shaving. But this is done at the cost of complex control techniques, changes in network operation and structure, high computational effort, large communication overhead, and complex fault protection. Moreover, the battery degradation is enhanced due to the increased number of charge/discharge cycles. As a result, the economic analysis of a particular charging strategy is essential. The bidirectional power flow strategies can be broadly classified into individual-based strategies and aggregator-based strategies [37], [38].

1) Individual-Based Strategies

These are very simple strategies that deal with each EV on an individual basis. When an EV is connected to the grid, the owner enters the final SoC and departure time. The load curve of the house and the electricity price curve are also made available. Such curves are based on measured or estimated values. The controller allocates the charging and discharging time slots based on the fact that the EV should be charged in low price hours and discharged at high price hours, provided that the EV is charged to the desired level at the end of the charging period and the SoC limitations of the battery are not violated [8], [39].

2) Aggregator-Based Strategies

The storage capacity of a single EV is very small from the grid's point of view. Using EVs individually for grid regulation is complex in terms of control, exhaustive in terms of communication, and less economical in terms of storage capacity and flexibility. So, many EVs are grouped and controlled as a whole. This is the essence of aggregator-based strategies [2], [40]. Some examples are as follows [40], [41].

Strategies Based on Load Frequency Control Signal

First of all, the current SoC of each EV is measured. Then the required SoC for the scheduled driving routine is estimated, keeping in view the charging routine, battery capacity, and system efficiency provided by the vehicle owner. If the required SoC is below the current SoC, it means the vehicle has surplus energy, and it can participate in V2G operation. Otherwise, the vehicle is to be charged [42].

In the second step, the participating power of the aggregator is determined by a multi-objective optimization problem to maximize the profits earned by V2G operation and minimize the tracking error of the load frequency control signal. The constraints are that the current SoC of each vehicle should not go below the SoC required for the driving demand during the up frequency regulation and above the maximum SoC limit during the down frequency regulation.

In the next step, the aggregator's participating power is allocated to each EV, which is to be charged or discharged. The objective is to minimize the change in the SoC of each EV under the constraints that the sum of individual vehicle allocated powers is equal to the participating power of the aggregator without exceeding the maximum charging/discharging power rating and without violating the SoC limitations of each battery [40], [43].

Strategies Based on Integration of Renewable Energy Sources

Probability density functions of driving and charging routines are determined based on available statistical data, and hence, a stochastic model for the power requirement of EVs is determined. Similarly, the available data for solar irradiance and wind speed, along with the respective plant capacities, help to model the output power of RESs. Network operators provide load and frequency regulation data on the basis of which

respective forecasts can be made. Frequency regulation data is an indication of the grid power requirement as a positive or negative reserve.

When a new EV arrives, its SoC is measured, and the owner is prompted to input the charging duration. After that, the grid power, power from RESs, and frequency regulation data are estimated. This data is fed to a controller, which determines the charging priorities. High priority means high charging power and vice versa. As different EVs have different arrival times, SoC and charging durations, different charging/discharging powers are assigned to these. For example, a vehicle with low initial SoC and a small charging duration requires high charging power, and it is unable to contribute to V2G operation. On the other hand, a vehicle with a high initial SoC and a long charging duration requires less charging power and can wait for off-peak and high renewable energy production times. It can be discharged during peak load times for grid contribution. Such vehicles are incentivized economically by dynamic pricing [**?**], [44], [45].

Strategies Based on Peak Load Reduction

Each registered EV owner is identified with a unique radio frequency identification tag. Whenever an authorized EV enters a charging station, the owner is prompted to specify its final SoC and departure time. The technical details, such as system efficiency, battery type, etc., can be extracted from the tag.

Based on this information, the charging time of the EV is estimated. If the charging time exceeds the departure time, the owner is prompted. The electricity price curve is fed to the controller, which is regularly updated based on available electricity market data. The price curve is quantized into a number of small intervals (usually 15 minutes) during which the price is assumed to be constant. Based on the charging time and electricity price, the cheapest time intervals are selected. In this way, the cheapest possible charging and peak load reduction are ensured.

If the owner allows for V2G operation, the time intervals with the highest price are determined for discharging under the constraint that the EV achieves its desired SoC at the moment of departure, and SoC limitations are not violated. Optimization of charging/discharging of EVs for the electricity price implicitly implies the optimization with respect to load demand [6], [46].

IV. STRATEGIES AT OPERATIONAL LEVEL

The above-mentioned control strategies involve EVs either at the component level or system level. Strategies can be developed at the operational level that can manage the charging of the EVs from a managerial point of view. A few are discussed below.

1) The discharged battery bank can be swapped with the charged one. This strategy adds enormous flexibility to EV scheduling but comes with cost, infrastructure, and regulation problems [14], [47].
2) The routes of EVs in a particular area are optimized and allocated efficiently to the available charging stations. This balances the load on charging stations and enables predictive modelling of charging behaviour. However, this approach is limited to a particular area and requires high computational effort for route modelling [48], [49].
3) EVs charged by an aggregator can be scheduled to share the energy stored in the batteries among themselves. EVs being charged in the homes can be used to provide electricity for the home during peak loads or faults, etc. This is called vehicle-to-home (V2H) operation [50].
4) Apart from the batteries, alternative energy storage systems, e.g., ultracapacitors and hydrogen-based energy storage systems, are under investigation [51], [52].
5) Apart from the physical connection for charging, electromagnetic phenomena can be used to charge the EVs in a wireless manner. This strategy has the advantages of safety and durability, but it has low efficiency and high power losses [53], [54].

V. CONCLUSIONS

The stress on fossil fuels has continuously increased over the past few decades, resulting in various environmental, economical and geopolitical problems. Electric vehicles can be used to reduce this stress if charged smartly. If not charged smartly, the vehicle electrification will be more harmful than beneficial. This paper discusses the strategies for smart charging of electric vehicles from the grid perspective. This means that the discussion on one hand is on the ways to reduce the burden on the power grid when an additional load of electric vehicles is added, and on the other hand, to use the energy storage capabilities of electric vehicles for grid support. As the first step, the selection of components for smart charging is discussed. Batteries, their charging methods, and chargers of different types are described. Then the strategies on the system level are discussed, which can be broadly classified into unidirectional and bidirectional power flow strategies. Unidirectional power flow strategies charge the electric vehicles from the grid, whereas the bidirectional power flow strategies not only charge the electric vehicles from the grid but also discharge them to support the grid when needed. Unidirectional power flow strategies are further classified based on centralized and decentralized strategies. Centralized strategies manage the charging of electric vehicles from a central control unit, whereas in decentralized strategies, the intelligence is distributed among the various components of the whole system, particularly the electric vehicles. The bidirectional power flow strategies can be split into individual and aggregator-based strategies. Individual-based strategies consider each electric vehicle on an individual level, whereas aggregator-based strategies consider a fleet of electric vehicles. Since the storage capacity of a single electric vehicle is small for the grid, the aggregator-based strategies are practically useful. In the end, some new ideas like battery swapping, route optimization, battery energy sharing, vehicle to home concept, usage of alternative energy storages and inductive charging are

discussed.

A next step can be to gather the research work done so far on these new ideas and to discuss their practical applicability. Various optimization techniques like genetic algorithm, particle swarm algorithm, interior point method, and bi-level programming, etc., which are usually used to implement these charging strategies, can be studied and compared as an extension of the present discussion. Moreover, the strategies outlined here can be used to improve the situation of electric vehicles in different case studies to bring pleasant effects for the grid integration of vehicle electrification.

ACKNOWLEDGMENT

The author is thankful to Dr.-Ing. Philipp Kuhn and Prof. Dr. rer. nat. Thomas Hamacher of the Chair of Renewable and Sustainable Energy Systems, Technical University of Munich, Germany, for their valuable suggestions for this work.

REFERENCES

[1] S. Deshmukh, H. Tariq, M. Amir, A. Iqbal, M. Marzband, and A. M. A. B. Al-Wahedi, "Impact assessment of electric vehicles integration and optimal charging schemes under uncertainty: A case study of qatar," IEEE Access, vol. 12, pp. 131350–131371, 2024.

[2] F. Mwasilu, J. J. Justo, E.-K. Kim, T. D. Do, and J.-W. Jung, "Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration," Renewable and Sustainable Energy Reviews, vol. 34, pp. 501 – 516, 2014.

[3] S. F. Tie and C. W. Tan, "A review of energy sources and energy management system in electric vehicles," Renewable and Sustainable Energy Reviews, vol. 20, pp. 82 – 102, 2013.

[4] Q. Wang, X. Liu, J. Du, and F. Kong, "Smart charging for electric vehicles: A survey from the algorithmic perspective," IEEE Communications Surveys Tutorials, vol. 18, pp. 1500–1517, Secondquarter 2016.

[5] S. Habib, M. Kamran, and U. Rashid, "Impact analysis of vehicle-togrid technology and charging strategies of electric vehicles on distribution networks – a review," Journal of Power Sources, vol. 277, pp. 205 – 214, 2015.

[6] S. Mal, A. Chattopadhyay, A. Yang, and R. Gadh, "Electric vehicle smart charging and vehicle-to-grid operation.," IJPEDS, vol. 28, no. 3, pp. 249– 265, 2013.

[7] M. Saklani, D. K. Saini, M. Yadav, and Y. C. Gupta, "Navigating the challenges of ev integration and demand-side management for india's sustainable ev growth," IEEE Access, vol. 12, pp. 143767–143796, 2024.

[8] J. C. Ferreira, V. Monteiro, J. L. Afonso, and A. Silva, "Smart electric vehicle charging system," in 2011 IEEE Intelligent Vehicles Symposium (IV), pp. 758–763, June 2011.

[9] R. Redwan, M. Hasan, A. Nadia, M. S. Khan, N. A. Chowdhury, and N.-U.-R. Chowdhury, "Design analysis and techno-economic assessment of a photovoltaic-fed electric vehicle charging station at dhaka-mawa expressway in bangladesh," Energy Conversion and Management: X, vol. 24, p. 100737, 2024.

[10] N. Bulawa, K. Mason, and F. Jacob, "Should the wheel be reinvented? market-referencing in the electric vehicle market charging infrastructure," Journal of Business Research, vol. 185, p. 114826, 2024.

[11] E. Xydas, C. Marmaras, and L. M. Cipcigan, "A multi-agent based scheduling algorithm for adaptive electric vehicles charging," Applied Energy, vol. 177, pp. 354 – 365, 2016.

[12] Z. Zhang and D. Gu, "Impacts of charging plug-in hybrid electric vehicles on the electric grid and its charging strategies," in 2012 Power Engineering and Automation Conference, pp. 1–4, Sept 2012.

[13] M. Amer, J. Masri, A. Dababat, U. Sajjad, and K. Hamid, "Electric vehicles: Battery technologies, charging standards, ai communications, challenges, and future directions," Energy Conversion and Management: X, vol. 24, p. 100751, 2024.

[14] J. Y. Yong, V. K. Ramachandaramurthy, K. M. Tan, and N. Mithulananthan, "A review on the state-of-the-art technologies of electric vehicle, its impacts and prospects," Renewable and Sustainable Energy Reviews, vol. 49, pp. 365 – 385, 2015.

[15] R. R. Kumar, C. Bharatiraja, K. Udhayakumar, S. Devakirubakaran, K. S. Sekar, and L. Mihet-Popa, "Advances in batteries, battery modeling, battery management system, battery thermal management, soc, soh, and charge/discharge characteristics in ev applications," IEEE Access, vol. 11, pp. 105761–105809, 2023.

[16] A. Upadhyaya and C. Mahanta, "An overview of battery based electric vehicle technologies with emphasis on energy sources, their configuration topologies and management strategies," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 2, pp. 1087–1111, 2024.

[17] K. Zhou, H. Yang, Y. Zhang, Y. Che, Y. Huang, and X. Li, "A review of the latest research on the topological structure and control strategies of onboard charging systems for electric vehicles," Journal of Energy Storage, vol. 97, p. 112820, 2024.

[18] J. García-Villalobos, I. Zamora, J. S. Martín, F. Asensio, and V. Aperribay, "Plug-in electric vehicles in electric distribution networks: A review of smart charging approaches," Renewable and Sustainable Energy Reviews, vol. 38, pp. 717 – 731, 2014.

[19] A. Di Giorgio, F. Liberati, and E. De Santis, "Electric vehicles charging: Technologies and use case driven control strategies," in Reference Module in Materials Science and Materials Engineering, Elsevier, 2024.

[20] H. Han, H. Xu, Z. Yuan, and Y. Zhao, "Interactive charging strategy of electric vehicles connected in smart grids," in Proceedings of The 7th International Power Electronics and Motion Control Conference, vol. 3, pp. 2099–2103, June 2012.

[21] H. Maghfiroh, O. Wahyunggoro, and A. I. Cahyadi, "Energy management in hybrid electric and hybrid energy storage system vehicles: A fuzzy logic controller review," IEEE Access, vol. 12, pp. 56097–56109, 2024.

[22] J. Ren, "Research on hybrid vehicle energy management strategy based on fuzzy control," in 2024 3rd International Conference on Energy and Power Engineering, Control Engineering (EPECE), pp. 127–131, 2024.

[23] S. Mehdi Rakhtala Rostami and Z. Al-Shibaany, "Intelligent energy management for full-active hybrid energy storage systems in electric vehicles using teaching–learning-based optimization in fuzzy logic algorithms," IEEE Access, vol. 12, pp. 67665–67680, 2024.

[24] A. A. Eajal, M. F. Shaaban, E. F. El-Saadany, and K. Ponnambalam, "Fuzzy logic-based charging strategy for electric vehicles plugged into a smart grid," in 2015 IEEE International Conference on Smart Energy Grid Engineering (SEGE), pp. 1–6, Aug 2015.

[25] Y. Shen, J. Xie, T. He, L. Yao, and Y. Xiao, "Ceemd-fuzzy control energy management of hybrid energy storage systems in electric vehicles," IEEE Transactions on Energy Conversion, vol. 39, no. 1, pp. 555–566, 2024.

[26] W. Jia, X. Pang, L. Shen, J. Tian, Y. Ma, and L. Liu, "Optimal scheduling method of electric vehicle charging and discharging considering peak cutting and valley filling," in 2024 36th Chinese Control and Decision Conference (CCDC), pp. 2940–2945, 2024.

[27] N. Attou, S.-A. Zidi, S. Hadjeri, and M. Khatir, "Improved peak shaving and valley filling using v2g technology in grid connected microgrid," in 2021 Third International Conference on Transportation and Smart Technologies (TST), pp. 53–58, 2021.

[28] G. Saren, Z. Sun, T. Liang, H. Li, P. Liu, D. Ma, H. Zhang, D. Sun, J. Yan, C. Li, and J. Li, "An optimized control strategy

for distributed energy storage system to reduce the peak-valley difference of distribution network," in 2021 IEEE 4th International Electrical and Energy Conference (CIEEC), pp. 1–5, 2021.

[29] L. Jian, Y. Zheng, and Z. Shao, "High efficient valley-filling strategy for centralized coordinated charging of large-scale electric vehicles," Applied Energy, vol. 186, Part 1, pp. 46 – 55, 2017.

[30] H. Turker, V. Pirsan, S. Bacha, D. Frey, J. Richer, and P. Lebrusq, "Heuristic strategy for smart charging of plug-in electric vehicle in residential areas: Variable charge power," in 2014 International Conference on Renewable Energy Research and Application (ICRERA), pp. 938–944, Oct 2014.

[31] M. Singh, A. P. Mittal, A. K. Singh, and K. Anand, "Heuristic power management of hybrid source electric vehicle based on pv-battery-pemfc," International Journal of Hydrogen Energy, vol. 80, pp. 475–485, 2024.

[32] S. S. Varghese, S. Q. Ali, and G. Joos, "Energy management of fast charging and ultra-fast charging stations with distributed energy resources," IEEE Access, vol. 12, pp. 131638–131655, 2024.

[33] L. Zhang, T. Chen, B. Yao, B. Yu, and Y. Wang, "Routing and charging scheduling for the electric carsharing system with mobile charging vehicles," Omega, p. 103211, 2024.

[34] I. S. Bayram, G. Michailidis, I. Papapanagiotou, and M. Devetsikiotis, "Decentralized control of electric vehicles in a network of fast charging stations," in 2013 IEEE Global Communications Conference (GLOBECOM), pp. 2785–2790, Dec 2013.

[35] M. Zulfiqar, Z. ul Abdeen, and M. Kamran, "Optimizing electric vehicle charging scheduling using enhanced multi-agent neural networks with dynamic pricing," Journal of Energy Storage, vol. 99, p. 113317, 2024.

[36] D. M. Glavic, "Agents and multi-agent systems: A short introduction for power engineers," May 2006.

[37] D. Choudhary, R. N. Mahanty, and N. Kumar, "Plug-in electric vehicle dynamic pricing strategies for bidirectional power flow in decentralized and centralized environment," Sustainable Energy, Grids and Networks, vol. 38, p. 101317, 2024.

[38] P. Vollmuth, D. Wohlschlager, L. Wasmeier, and T. Kern, "Prospects of electric vehicle v2g multi-use: Profitability and ghg emissions for use case combinations of smart and bidirectional charging today and 2030," Applied Energy, vol. 371, p. 123679, 2024.

[39] M. Su, S. Wu, H. Dan, J. Xu, Y. Sun, H. Wang, Y. Liu, W. Xiong, and X. Liang, "A natural bidirectional isolated single-phase ac/dc converter with wide output voltage range for aging test application in electric vehicle," IEEE Journal of Emerging and Selected Topics in Power Electronics, vol. 9, no. 3, pp. 3489–3500, 2021.

[40] C. Peng, J. Zou, L. Lian, and L. Li, "An optimal dispatching strategy for {V2G} aggregator participating in supplementary frequency regulation considering {EV} driving demand and aggregator's benefits," Applied Energy, vol. 190, pp. 591 – 599, 2017.

[41] S. Hussain, A. P. Teni, I. Hussain, Z. Hussain, F. Pallonetto, J. Eichman, R. R. Irshad, I. M. Alwayle, M. Alharby, M. A. Hussain, M. F. Zia, and Y.- S. Kim, "Enhancing electric vehicle charging efficiency at the aggregator level: A deep-weighted ensemble model for wholesale electricity price forecasting," Energy, vol. 308, p. 132823, 2024.

[42] A. Kazemtarghi, S. Dey, and A. Mallik, "Optimal utilization of bidirectional evs for grid frequency support in power systems," IEEE Transactions on Power Delivery, vol. 38, no. 2, pp. 998–1010, 2023.

[43] T. Liu, P. Wang, Q. Peng, M. Zhang, T. Wang, and J. Meng, "Operationarea-constrained adaptive primary frequency support strategy for electric vehicle clusters," Journal of Modern Power Systems and Clean Energy, vol. 11, no. 6, pp. 1982–1994, 2023.

[44] M. Abdelsattar, M. A. Ismeil, M. M. Aly, and S. Saber Abu-Elwfa, "Analysis of renewable energy sources and electrical vehicles integration into microgrid," IEEE Access, vol. 12, pp. 66822–66832, 2024.

[45] T. Ma and O. A. Mohammed, "Economic analysis of real-time large-scale pevs network power flow control algorithm with the consideration of v2g services," IEEE Transactions on Industry Applications, vol. 50, pp. 4272– 4280, Nov 2014.

[46] A. Srivastava, M. Manas, and R. K. Dubey, "Integration of power systems with electric vehicles: A comprehensive review of impact on power quality and relevant enhancements," Electric Power Systems Research, vol. 234, p. 110572, 2024.

[47] B. V. Vani, D. Kishan, M. W. Ahmad, and B. N. K. Reddy, "An efficient battery swapping and charging mechanism for electric vehicles using bat algorithm," Computers and Electrical Engineering, vol. 118, p. 109357, 2024.

[48] Q. R. Zhang, J. Wang, C. Wang, and S. Ren, "Research on electric vehicle route planning and energy consumption prediction based on CNN-LSTM model," in 2024 12th International Conference on Traffic and Logistic Engineering (ICTLE), pp. 184–188, 2024.

[49] M. Blasio and F. Cipparrone, "Sustainable paths: Strategies for efficient routes in reducing fuel consumption and pollutant emissions in são paulo," in 2024 Mediterranean Smart Cities Conference (MSCC), pp. 1–4, 2024.

[50] G. Kumar and S. Mikkili, "Critical review of vehicle-to-everything (v2x) topologies: Communication, power flow characteristics, challenges, and opportunities," CPSS Transactions on Power Electronics and Applications, vol. 9, no. 1, pp. 10–26, 2024.

[51] R. Jasper, A. Luxa, and G. Lichtenberg, "Influence of regulation on the operation of a hydrogen-based energy storage system including economic model predictive control: Use case of a German subsidy scheme," IEEE Access, vol. 12, pp. 146686–146697, 2024.

[52] Z. Amjadi and S. S. Williamson, "Review of alternate energy storage systems for hybrid electric vehicles," in 2009 IEEE Electrical Power Energy Conference (EPEC), pp. 1–7, Oct 2009.

[53] V. Ramadoss, B. Chandrasekar, M. Ahmed, D. Savio A, N. Rajamanickam, and T. A. H. Alghamdi, "Research insights on recent power converter topologies and control strategies for wireless ev chargers: A comprehensive study," IEEE Open Journal of Power Electronics, vol. 5, pp. 1641– 1658, 2024.

[54] H. Shareef, M. M. Islam, and A. Mohamed, "A review of the state-of-the-art charging technologies, placement methodologies, and impacts of electric vehicles," Renewable and Sustainable Energy Reviews, vol. 64,pp 403 – 420, 2016.

Date available online: 09-01-2026
Vol. 3, Issue 2 (July – December 2025)

DOI: https://doi.org/10.24312/ucp-jeit.03.02.573

Multi-Class Brain Tumor Detection Using Transfer Learning and Interpretable Deep Models

Muhammad Asif Feroz[1], Anam Safdar Awan[2], Fareeha Batool[3], Narges Shahbaz[4], Anam Murtaza[5], Kamran Ali[6]

[1]Department of Computer Science and IT, Superior University, Lahore, 54000, Pakistan (e-mail: asif.janjua005@gmail.com)
[2]Department of Computer Science and IT, Superior University, Lahore, 54000, Pakistan (e-mail: anamawan416@gmail.com)
[3]Department of CS & IT, Superior University, Sargodha Campus, Pakistan (e-mail: fareehabaloch29@gmail.com)
[4]Department of Arts and Humanities, Superior University, Lahore, Pakistan (e-mail: nargesshahbaz20137@gmail.com)
[5]Department of Biology, Superior University, Sargodha Campus, Pakistan (e-mail: anammurtazashah434@gmail.com)
[6]Department of CS & IT, University of Sargodha, Sargodha, Pakistan (e-mail: kamranali3037414518@gmail.com)

Corresponding author: Muhammad Asif Feroz (e-mail: asif.janjua005@gmail.com).

ABSTRACT

Accurate brain tumor detection remains critical yet challenging due to diagnostic complexity and variability in MRI interpretation. This study proposes a deep learning approach for automated multi-class brain tumor classification using transfer learning (TL). Three pre-trained CNN models, ResNet50, InceptionV3, and VGG16, were adapted and evaluated on a curated MRI dataset of 7,000+ images. Preprocessing, feature extraction, fine-tuning, and integration of Explainable AI (Grad-CAM, LIME, SHAP) ensured robust and interpretable results. ResNet50 achieved the highest performance with 98% accuracy, 0.92 F1-score, and 0.96 AUC, outperforming the other models across all metrics, with strong convergence and minimal misclassification. ResNet50's architecture enabled deeper feature learning and improved generalization. Explainable AI visualizations confirmed model focus on tumor-relevant MRI regions, enhancing clinical interpretability. The findings position ResNet50 as an effective and explainable solution for MRI-based brain tumor classification, suitable for future real-world deployment and further expansion to mobile and multi-center applications.

INDEX TERMS: Brain Tumor Detection, Deep Learning, ResNet50, MRI Classification, CNN, Medical Imaging, Binary Classification, Tumor Diagnosis.

I. INTRODUCTION

Brain tumors are among the most critical neurological disorders, characterized by aberrant development of cells inside or around the brain that perturb normal brain function [1]. Depending on their nature, brain tumors are often classed into *benign* (non-cancerous and slow-growing) and *malignant* (cancerous and aggressive) [2]. According to the International Agency for Research on Cancer (IARC), more than 126,000 new brain tumor cases are diagnosed annually worldwide, with over 97,000 deaths attributed to the disease each year. The World Health Organization (WHO) further projects a 5% annual increase in brain tumor cases globally, making early detection and effective treatment increasingly vital [3; 4].

The early and correct diagnosis of brain tumors plays a vital role in enhancing patient outcomes, reducing mortality rates, and planning personalized treatment strategies [5]. Magnetic Resonance Imaging (MRI) has evolved as a key imaging technique owing to its non-invasive quality and ability to obtain superior resolution soft-tissue contrasts. However, the manual interpretation of MRI images by radiologists is a time-consuming process that is susceptible to diagnostic inconsistencies, inter-observer variability, and potential oversight, especially when dealing with large imaging datasets [6; 7].

In the past couple of years, the rise of Artificial Intelligence (AI), notably deep learning (DL), has transformed the landscape of medical image analysis [8]. AI models are now capable of learning complex, non-linear representations from raw image data, thus assisting healthcare professionals in decision-making processes [9]. Among these models, Convolutional Neural Networks (CNNs) have shown remarkable performance in a broad variety of computer vision applications, including classification of images, segmentation of images, and object recognition in images. Their success in the biomedical domain has led to promising outcomes in brain tumor classification, localization, and segmentation [10].

However, deep CNNs require substantial labeled data and computational resources for training from scratch, which poses a significant limitation in medical imaging, where curated and annotated datasets are limited due to privacy concerns, expert availability, and patient variability. To overcome this challenge, transfer learning has emerged as a strong alternative. Transfer learning facilitates the use of pretrained deep learning (DL) models, generally developed on massive datasets like ImageNet, to be fine-tuned or adapted for specific medical tasks with minimal training data and computational cost.

Recent investigations have examined the use of CNN models such as VGG-16, ResNet, Inception, and hybrid models for the classification of brain tumors. For example, InceptionV3 was used with ensemble classifiers to achieve high accuracy on brain MRI scans. A CNN-SVM combination was employed and achieved over 95% classification accuracy. The fine-tuned versions of VGG and ResNet were used to improve performance. These studies confirm the viability of DL-based methods but often focus on binary classification (tumor vs. no tumor) or evaluate a single model architecture in isolation [11; 12].

Moreover, there are several remaining limitations in existing literature:

- Lack of comparative analysis across multiple pretrained CNN architectures using a consistent dataset and evaluation framework.
- Absence of multi-class classification studies that distinguish between glioma, meningioma, pituitary tumor, and no tumor categories, which is essential for real-world clinical deployment.
- Minimal investigation into the impact of different transfer learning strategies (i.e., feature extraction vs. finetuning) under the same experimental setup.
- Limited exploration of resource-efficient models suitable for deployment in hospitals with constrained computing environments.

A. MOTIVATION AND OBJECTIVE

This work intends to avert the gaps by setting up a transfer learning-based deep learning framework that applies and compares three state-of-the-art pre-trained CNN architectures, ResNet-50, InceptionV3, and VGG-16 for multi-class brain tumor classification. The models are evaluated using a comprehensive MRI dataset obtained from multiple open-access sources, including Br35H, SARTAJ, and Figshare, containing over 7,000 labeled images. Both feature extraction and fine-tuning strategies are employed to investigate the effect of transfer learning depth on classification performance.

Through extensive experimentation and evaluation using metrics such as F1-score, accuracy, precision, and recall, the research seeks to discover the optimum model configuration for real-world deployment. The overarching goal is to build an automated, accurate, and resource-efficient computer-aided diagnostic (CAD) mechanism for the prompt identification and categorization of brain tumors, thereby reducing radiologists' workload and enhancing diagnostic confidence in clinical environments.

B. KEY CONTRIBUTIONS

The significant advancements of this work are outlined as follows:

1) **Development of a Transfer Learning Framework:** A robust and scalable deep learning system is provided for diagnosing brain cancers from MRI images, utilizing transfer learning on pre-trained CNN architectures, ResNet-50, InceptionV3, and VGG-16.

2) **Multi-Class Brain Tumor Classification:** The study addresses a four-class classification problem involving gliomas brain tumor, meningiomas brain tumor, pituitary tumors, and no brain tumor categories. This enhances the clinical relevance of the proposed system beyond binary classification.

3) **Comparative Analysis of Transfer Learning Strategies:** Both feature extraction and fine-tuning techniques are implemented and analyzed under the same experimental settings to assess their effectiveness on medical image classification tasks.

4) **Utilization of a Large and Diverse Dataset:** A comprehensive brain MRI dataset comprising over 7,000 labeled images from multiple publicly available sources (Br35H, SARTAJ, and Fig-share) is curated and used for training, validation, and testing.

5) **Performance Evaluation Using Multiple Metrics:** The models are tested using important classification metrics that involve precision, recall, F1-score, and accuracy, along with confusion matrices for detailed performance assessment.

6) **Design of a Resource-Efficient AI Solution:** The research demonstrates that high-performance classification can be achieved without training models from scratch, making the proposed solution viable for deployment in resource-constrained clinical environments.

7) **Key feature identification using XAI:** To improve the transparency and interpretability of the model, the Explainable AI (XAI) Grad-CAM model is used, which has generated heat maps of MRI images that highlight the regions of the brain tumor.

The rest of this work is organized as follows: Section II analyzes relevant work on brain tumor detection using machine learning and deep learning, noting gaps in the field. Section III explains the suggested technique, including dataset preparation, transfer learning methodologies (feature extraction vs. fine-tuning), and model architectures (ResNet50, InceptionV3, VGG16). Section IV describes the experimental setup, evaluation metrics, and hardware configuration. Section V presents the results, with a comparative analysis of model performance across accuracy, loss, F1-score, and AUC. Finally, Section VI concludes the study, discusses clinical implications, and suggests future directions.

II. RELATED WORK

Over the last decade, the integration of Artificial Intelligence (AI) technologies, notably Machine Learning (ML) and Deep Learning (DL), into medical imaging has significantly transformed brain tumor detection and classification methodologies. These innovations have brought promising advancements in terms of diagnostic automation, accuracy, and efficiency. However, despite the increasing adoption of AI in neuroimaging, several

persistent limitations in current research impede widespread clinical adoption, including poor model generalizability, lack of interpretability, computational inefficiencies, and limited scalability to real-world medical environments.

Initial approaches in this domain largely relied on traditional machine learning methods applied to handcrafted features extracted from MRI images. For instance, the study *Design and Analysis for Advancements in Brain Tumor Detection Model by using Machine Learning (ML) Techniques* employed classical ML algorithms to process and classify MRI scans. Although these models demonstrated a baseline capacity to differentiate between tumor types, they struggled with low segmentation precision, high false positive rates, and poor adaptability across datasets with different acquisition parameters [12]. These limitations underscore the challenges posed by manual feature engineering and rule-based classification strategies in complex imaging tasks.

To systematically assess developments in this domain, several literature reviews and meta-analyses have been conducted. The *Systematic Literature Review on ML and DL from 2013 to 2023* compiles a decade's worth of studies and reveals a heavy dependence on annotated datasets, inconsistent imaging protocols, and a lack of standardized evaluation benchmarks. These issues severely limit model reproducibility and generalization, particularly when deployed across diverse clinical institutions [13].

Moreover, survey-based studies such as *Brain Tumor Identification and Classification Using Machine Learning (ML): An In-Depth Survey* and *Brain Tumor Identification Using Machine Learning (ML)* have highlighted the evolution of ML in neuro-oncology while identifying several inherent limitations. These include high intra-class variance due to morphological differences among tumor types, inter-scanner variability, and the time-consuming nature of manual diagnostic processes. These studies emphasize that traditional ML systems are often error-prone and inefficient for real-time decision-making, particularly in resource-limited clinical environments [14; 15].

The transition to deep learning marked a significant leap in model performance, particularly for feature extraction and classification by using convolutional neural networks (CNNs). Nonetheless, several DL-based studies exhibit shortcomings. For instance, works such as *Classification of Brain Tumor Detection Techniques: A Review* and *Empowering Healthcare with AI* introduced hybrid AI approaches that combine multiple ML/DL models. While these architectures yielded improved accuracy, they suffered from increased model complexity, higher computational costs, and difficulty in deployment due to hardware constraints and the requirement for large annotated datasets [16; 17].

Other integrative studies, like *Brain Tumor Detection: Integrating ML and DL*, explored dual-pipeline systems combining traditional ML with CNN-based classifiers. Although these attempts sought to utilize the best of both worlds, they resulted in increased training duration, limited scalability, and suboptimal performance when applied to multi-class tumor classification scenarios [18].

Deep learning architectures specifically designed for medical image analysis, such as InceptionV3 and ResNet-50, have demonstrated state-of-the-art results in tumor classification tasks. For example, the work *An Inception V3-Based Glioma Brain Tumor Detection in MRI Images* leveraged deep CNNs for detecting gliomas with high accuracy. However, the model required extensive hyperparameter tuning and access to high-quality, annotated data, making it unsuitable for deployment in low-resource hospitals [16]. Similarly, *Deep Learning-Enhanced MRI for Brain Tumor Detection* showcased improved feature learning through DL but faced overfitting issues due to limited sample diversity and a lack of interpretability mechanisms [19].

Further, the study *Optimizing Brain Tumor Classification with ResNet-50 Feature Extraction* examined the effectiveness of residual networks in extracting hierarchical features from MRI data. Despite achieving impressive accuracy metrics, the computational demands of ResNet-50 present a practical barrier to its clinical application, particularly in rural or under-resourced settings [6].

Comparative analyses, such as *A Comparative Study of DL vs. ML*, clearly show the superiority of deep learning in terms of raw performance but also expose concerns regarding training time, memory consumption, and lack of transparency in decision-making processes. These limitations hinder clinical trust and adoption, especially when models are treated as black-box systems [17].

The problem of data imbalance and generalization is also prominent in studies like *Identification of Challenges and Limitations in Detection and Segmentation of Brain Tumors*. These works identify key challenges, including skewed class distributions (e.g., more glioma cases than meningioma), segmentation inaccuracies, and a lack of robust evaluation frameworks that cover both tumor detection and multi-class classification [20].

To improve upon traditional and deep learning approaches, hybrid models have also been introduced. *Brain Tumor Detection Using Hybrid Machine Learning Models* proposes an ensemble-based ML approach to enhance predictive accuracy. While performance improvements were noted, the added complexity and extended training requirements complicate clinical deployment timelines and maintenance cycles [18].

From this extensive literature review, it becomes

TABLE 1: Comparison of Existing Work vs. Proposed Methodology

Study / Reference	Limitations Identified	Our Proposed Solution
Design and Analysis for Advancements in Brain Tumor Detection [12]	Low segmentation accuracy; high false positives	ResNet50, InceptionV3, and VGG-16 with robust feature learning and reduced false detection.
Systematic Literature Review on ML and DL (2013–2023) [13]	Heavy reliance on labeled data; inconsistent quality	Transfer learning with pre-trained models lowers annotation dependency
Brain Tumor Detection Using Machine Learning: A Comprehensive Survey [14]	Morphological variation and imaging inconsistency	Multi-class classification across four tumor types improves generalizability
Brain Tumour Detection Using Machine Learning [15]	Time-consuming manual analysis; prone to error	End-to-end automated deep learning classification
Classification of Brain Tumor Detection Techniques: A Review [16]	Tumor variability impacts detection accuracy	CNN models trained on diverse and augmented datasets
Empowering Healthcare with AI [17]	Limited annotated MRI data; overfitting risk	Combines large public datasets with augmentation and regularization
Brain Tumor Detection: Integrating ML and DL [18]	Complex model integration and training duration	Lightweight architecture and efficient transfer learning strategies
Deep Learning-Enhanced MRI for Brain Tumor Detection [19]	Overfitting on small datasets; poor interpretability	Standardized dataset and benchmarking; visual explainability planned (e.g., Grad-CAM)
An InceptionV3-Based Glioma Detection [16]	Requires large annotated data and hyperparameter tuning	Efficient use of public datasets with less tuning via transfer learning
Optimizing Brain Tumor Classification with ResNet-50 [6]	Computational cost limits deployment	Balanced performance and efficiency in clinical settings using fine-tuning
A Comparative Study of DL vs. ML [17]	DL models not feasible for low-resource clinics	Designed for high accuracy and low hardware requirements
Identification of Challenges in Tumor Segmentation [20]	Class imbalance and segmentation errors	Balanced multi-class dataset and evaluation metrics used
Brain Tumor Detection Using Hybrid ML Models [18]	High complexity and extended training time	Streamlined transfer learning framework for quick deployment

Evident that most existing studies focus on binary classification (tumor vs. no tumor), evaluate a single network architecture in isolation, or fail to investigate different transfer learning strategies comprehensively. More critically, very few works address the problem of computational feasibility in real-world clinical workflows, especially those involving high-resolution images and multi-class tumor scenarios [19; 20]. To address these gaps, our proposed research introduces a robust and unified transfer learning framework that:

- Performs four-class categorization spanning gliomas, meningiomas, pituitary tumor and no tumor categories.
- Evaluates and compares three state-of-the-art pretrained CNN architectures: ResNet-50, InceptionV3, and VGG-16.
- Benchmarks two core transfer learning strategies, feature extraction and fine-tuning, under a uniform experimental protocol.
- Emphasizes computational efficiency, thereby enabling practical deployment in both well-equipped and resource-constrained clinical environments.

This work aims not only to improve classification performance but also to bridge the translational gap between model development and clinical application. Our approach incorporates real-world constraints and focuses on generalizability, interpretability, and scalability to ensure relevance and impact in actual diagnostic settings, as shown in Table I.

III. SYSTEM METHODOLOGY

This section elaborates on the full technique utilized for the creation of a transfer learning-based brain tumor classification system employing deep convolutional neural networks (CNNs). The proposed methodology tries to solve the shortcomings mentioned in previous techniques by using the capabilities of three pre-trained models, ResNet-50, InceptionV3, and VGG-16 on a large-scale, multi-class MRI dataset. The methodology is composed of several stages: data acquisition and preprocessing, architecture adaptation, transfer learning strategy, training algorithms, optimization methods, performance evaluation, and integration of Explainable AI (XAI) for model interpretability.

A. OVERVIEW OF THE PROPOSED FRAMEWORK

The proposed system consists of an end-to-end deep learning pipeline designed for the classification of brain MRI images into four diagnostic categories: gliomas, meningiomas, pituitary tumors, and no tumor. Each input image undergoes a standardized preprocessing phase before being passed into one of the selected pre-trained CNN architectures. The models are adapted

using transfer learning techniques, either feature extraction or fine-tuning, to classify tumors effectively with limited training data. To ensure trust and clinical acceptance, Explainable AI (XAI) approaches are incorporated to bring visibility into the model's process of decision-making.

B. DATASET DESCRIPTION AND PREPROCESSING

The MRI dataset employed in this study comprises a combination of three publicly available sources: Br35H, SARTAJ, and the Fig share repository. These datasets contain axial T1-weighted contrast-enhanced (T1W-CE) MRI brain images with corresponding annotations across four categories. In total, 7,022 images were collected and layered into training (70%), validation (15%), and test (15%) subsets to ensure a balanced evaluation.

Image preprocessing is critical for standardizing data input across different models and includes the following steps:

- **Resizing:** Images are resized to 224×224 pixels for ResNet-50 and VGG-16, and 299×299 pixels for InceptionV3 to match the input layer specifications.
- **Normalization:** Pixel intensity data is adjusted to the [0, 1] range to ensure uniform input.
- **Data Augmentation:** Techniques such as horizontal/vertical flips, zooming, and random rotations are used to artificially increase the dataset and enhance generalization.
- **Label Encoding:** Class labels are single-hot encoded to meet the categorical output format of the models.

C. TRANSFER LEARNING STRATEGY

Transfer learning is leveraged to reuse knowledge acquired from models trained on the ImageNet dataset. Two approaches are employed:

1) **Feature Extraction:** The pre-trained convolutional base is frozen, and only the top classification layers are retrained on the MRI dataset, as shown in Figure 1. This method is computationally efficient and less prone to overfitting.

2) **Fine-Tuning:** A portion of the higher-level convolutional layers is unfrozen and retrained alongside the classifier. This allows the model to learn domain-specific features relevant to MRI data, offering better performance when the training data is moderately sized.

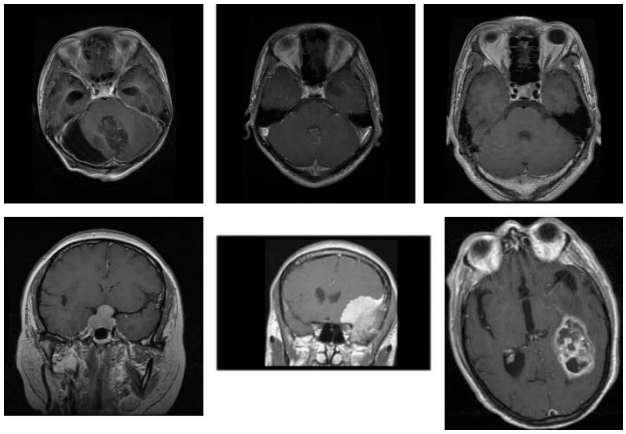

FIGURE 1: Samples of datasets used in training and testing.

D. MODEL ARCHITECTURE ADAPTATION

In this study, three widely recognized pre-trained

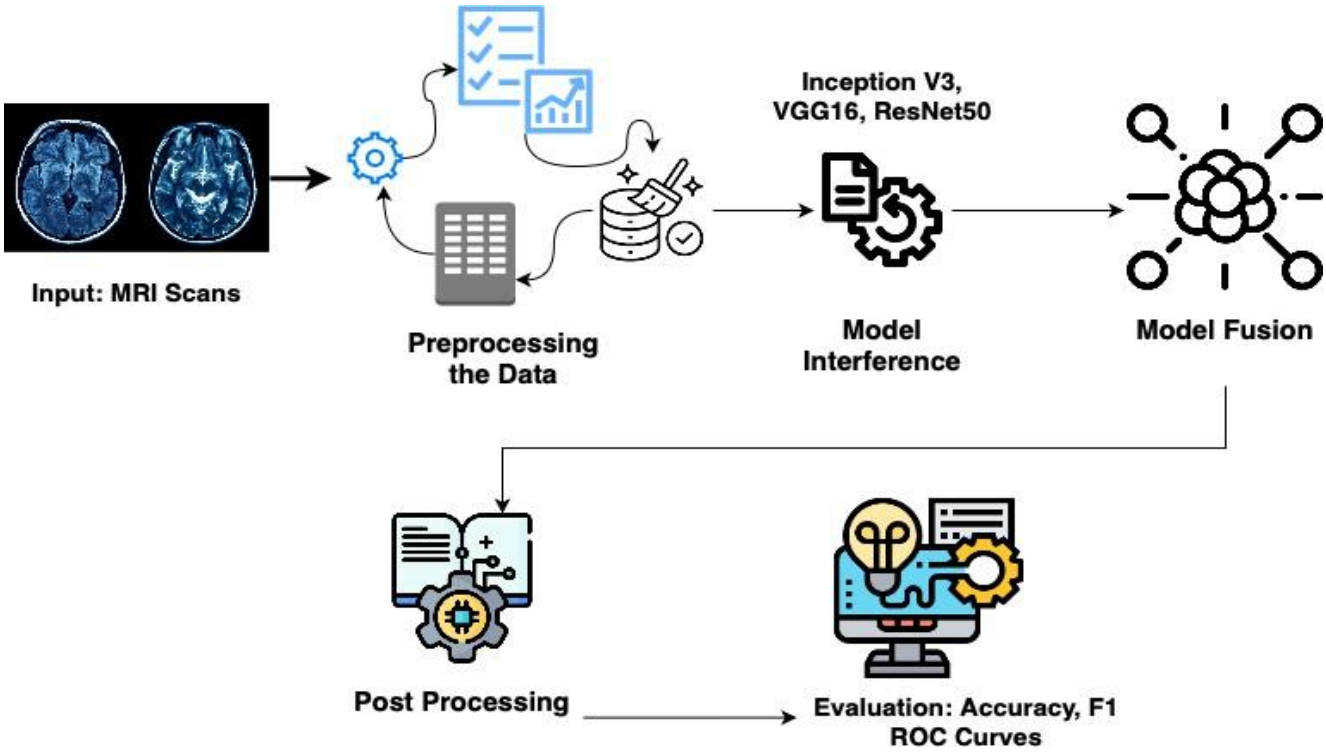

FIGURE 2: Adapted architectures of VGG-16, ResNet-50, and InceptionV3 with transfer learning classifier heads.

Convolutional Neural Network (CNN) architectures, Res-Net-50, VGG-16, and InceptionV3, are adapted to perform multiclass classification of brain tumors. These designs are chosen for their shown efficacy in large-scale visual identification challenges and their capacity to generalize to medical imaging domains via transfer learning, as seen in Figure 2.

The VGG-16 architecture is a 16-layer deep CNN that employs a simple and consistent design pattern of stacked three-by-three 3x3 convolutional layers, succeeded by max-pooling layers. It is known for its depth and uniform structure, which makes it both interpretable and effective for transfer learning. In this work, the original classifier head of VGG-16 is removed and replaced with a custom classification block consisting of a Flatten layer, a fully connected A dense layer comprising 512 units with ReLU activation, followed by a Dropout layer (rate 0.5) to mitigate overfitting, and concluding with a final dense layer including 4 output neurons and softmax activation to support multi-class prediction.

The ResNet-50 deep residual network model is a 50-layer network that introduces identity-based skip connections, allowing gradients to bypass one or more layers during backpropagation. This solution directly tackles the vanishing gradient issue that often impacts deep neural networks. By facilitating the training of far deeper structures, ResNet-50 can capture complex and abstract features within MRI data. In this framework, the final fully connected layers of ResNet-50 are replaced with a Global Average Pooling (GAP) layer that follows a Dense classification layer with softmax activation to produce class probabilities for the four tumor kinds.

The InceptionV3 architecture is a highly modular CNN that utilizes inception modules, which perform multiple convolution operations in parallel (e.g., 1×1, 3×3, 5×5) within the same layer. This design enhances

E. TRAINING ALGORITHMS AND OPTIMIZATION

Two training algorithms are proposed to guide the model learning process:

Algorithm 1: Baseline Transfer Learning Classifier (21) This algorithm initializes the pre-trained CNN with its convolutional base frozen (feature extraction), appends a custom classifier head, and trains only the added layers using categorical cross-entropy.

Algorithm 2: Progressive Fine-Tuning Strategy (22) This advanced strategy begins by training the classifier head (as in Algorithm 1), then progressively unfreezes deeper layers of the convolutional base for additional training. A small learning rate is maintained to avoid destabilizing pretrained weights. This staged unfreezing allows gradual domain adaptation.

Optimization: All models are built on the Adam optimizer with a learning rate of $\eta = 10^{-4}$, categorical cross-entropy loss, and accuracy as the main performance indicator. Regularization methods such as dropout and early halting are applied to avoid overfitting.

Algorithm 1 Baseline Transfer Learning Classifier

Require: Pre-trained CNN f_θ, dataset D, batch size B, number of epochs N
Ensure: Trained model f_θ
Freeze all convolutional layers of f_θ
Append custom classifier head to f_θ
for epoch $e = 1$ to N **do**
  for each batch (x, y) in D **do**
    $\hat{y} \leftarrow f_\theta(x)$ ▷ Forward pass
    Compute loss $\mathcal{L}(\hat{y}, y)$
    Update classifier head weights using backpropagation
  end for
end for
return f_θ ▷ Fine-tuned model

Algorithm 2 Progressive Fine-Tuning Strategy

Require: Pre-trained CNN f_θ, dataset D, number of epochs N, unfreeze depth d, batch size B
Ensure: Fine-tuned model f_θ
Freeze all layers of f_θ; append classification head
Train classifier head on D for a few initial epochs
Unfreeze the top d layers of the CNN base
Re-compile the model with a reduced learning rate $\eta \ll 1$
for epoch $e = 1$ to N **do**
  for each batch (x, y) in D **do**
    Perform forward pass and compute predictions $\hat{y} \leftarrow f_\theta(x)$
    Compute loss $\mathcal{L}(\hat{y}, y)$
    Backpropagate and update weights using Adam optimizer
  end for
end for
return f_θ ▷ Fine-tuned model

F. EXPLAINABLE AI INTEGRATION

To boost model transparency and interpretability, there are several Explainable AI (XAI) strategies, such as Grad-CAM, LIME, and SHAP, offering insights into the decision-making process of the trained models. The description of each strategy is given below. As our dataset is large, LIME and SHAP are computationally expensive to handle such a large dataset, Grad-CAM the XAI method, is applied in this framework.

- **Grad-CAM (Gradient-weighted Class Activation Mapping):** This approach provides heatmaps that show the areas of the MRI images most relevant in the model's decision-making process. By visualizing the importance of specific areas of the brain in relation to tumor classification, Grad-CAM helps clinicians understand what parts of the MRI image the model focuses on.
- **LIME (Local Interpretable Model-agnostic Explanations):** LIME predicts the model's behavior locally, producing interpretable explanations for individual predictions. This can help explain why the model classified an image into a particular tumor

category, providing insights into specific features that drove the classification.

- **SHAP (Shapley Additive Explanations):** SHAP values decompose the model's prediction into the contribution of each feature (e.g., pixel region) to the final classification. This global explanation technique helps quantify the importance of various image regions across the entire dataset.

Grad-CAM XAI method is incorporated into the model evaluation phase, where it provides visual and numerical explanations of the model's reasoning for each prediction, enhancing trust and transparency.

G. EVALUATION METRICS

The proposed models are evaluated using:

- **Accuracy:** Proportion of correctly categorized positive and negative instances on brain magnetic resonance images.
- **Precision:** Correct True Positive (TP) predictions per total predicted True Positive (TP).
- **Recall:** Correct True Positive (TP) predictions per actual True Positive (TP).
- **F1-Score:** Harmonic mean of precision metrics and recall metrics.
- **Confusion Matrix:** A visual matrix of true labels vs predicted labels.
- **XAI Explanation Consistency:** Analysis of the consistency and reliability of explanations across different model runs.

These metrics ensure a comprehensive evaluation across all tumor classes and classification challenges. The integration of XAI enables robust performance benchmarking, model adaptability for real-world clinical deployment, and enhanced interpretability, essential for gaining clinical acceptance and ensuring patient safety.

IV. EXPERIMENTAL SETUP

In this study, we evaluate four deep learning (DL) models, **CNN**, **VGG16**, **ResNet50**, and **InceptionV3**, for brain tumor detection using a publicly available MRI brain tumor dataset. Below are the details of the experimental setup, including dataset description, model training, and evaluation procedure.

A. DATASET DESCRIPTION

The dataset employed in this work is the **Brain MRI images Dataset**, which comprises tagged images of brain MRIs, with two primary classes: tumor and non-tumor. The collection comprises pictures of varied resolutions and kinds of MRI scans, e.g., T1-weighted (T1W), T2-weighted (T2W). The total count of images is 7022. The photos were separated into training (80%) and testing (20%) groups to make sure that the dataset was balanced across the classes.

B. PREPROCESSING

Before feeding the MRI images into the models, several preprocessing steps were performed:

- **Resizing**: All the images were scaled to 224x224 pixels to satisfy the input needs of the models.
- **Normalization**: The pixel values of the images were standardized to the range [0, 1] to accelerate the training process and increase convergence.
- **Augmentation**: To strengthen the durability of the models and minimize overfitting, data augmentation methods such as random rotation, flipping, and zooming were added to the training set.

C. MODEL ARCHITECTURE AND HYPERPARAMETERS

A basic convolutional neural network (CNN) model consists of 3 convolutional layers, followed by max-pooling, and a fully connected layer for classification, which provides the basis for different models used in the framework. Resnet-50, VGG-16, and Inception-V3 architectures of CNN were used for the comparison:

- **VGG16**: A deeper network with 16 layers comprised of convolutional layers followed by fully linked layers. Pre-trained weights from ImageNet were utilized to fine-tune the model (23).
- **ResNet50**: A residual network with 50 layers was designed to handle the issue of the vanishing gradient. It includes skip connections to allow deeper models to be trained (24).
- **InceptionV3**: A model designed by Google for image classification. It uses auxiliary classifiers and factorized convolutions, which make it more efficient in terms of both speed and accuracy (25).

The following hyperparameters were used across all models:

- **Learning Rate**: 0.0001 for all models.
- **Batch Size**: 32 images are in a single batch.
- **Epochs**: 50 epochs were used to train the model.
- **Optimizer**: Adam optimizer with a learning rate decay of 0.9.
- **Loss Function**: Categorical Cross-Entropy was employed as the loss function for multiclass classification.

D. TRAINING AND EVALUATION

Each model was trained on the training set and evaluated on the testing set using several performance metrics:

- **Accuracy:** The proportion of correctly categorized brain magnetic resonance images.
- **Precision:** Correct predictions of true positive (TP) per total of predicted true positives (TP) and False negative (FN).
- **Recall:** The correct predictions of True Positive (TP) per actual True Positive (TP).
- **F1-Score:** Harmonic mean of precision metrics and recall metrics.
- **Area Under the ROC Curve (AUC)**: A measure of the model's ability to discriminate between the classes.

The training was performed on a machine with an NVIDIA Tesla V100 GPU, which accelerated the training of the models.

E. EVALUATION METRICS AND VALIDATION

The models were evaluated using the following:

- **Confusion Matrix**: To understand the distribution of True Positives (TP), False Positives (FP), True

Negatives (TN), and False Negatives (FN) for each model.

- **ROC Curve**: A graphical representation of the true positive rate against the false positive rate, used to visualize the performance across different thresholds.
- **Loss Curve**: To observe the convergence behavior and the extent of overfitting during training.
- **Precision-Recall Curve**: To assess the balance between precision and recall, especially in cases of imbalanced datasets.

F. HARDWARE SETUP

To effectively train and evaluate the proposed deep learning models, a high-performance computational environment was utilized. The hardware specifications of the experimental setup are outlined below:

- **CPU**: Intel Core i9-10900K, 10-core processor clocked at 3.7 GHz, providing exceptional single-thread and multi-thread performance suitable for parallel data preprocessing and I/O operations.
- **GPU**: NVIDIA Tesla V100, equipped with 32GB of VRAM, is a contemporary accelerator suitable for deep learning tasks. The Tensor cores significantly improve matrix multiplication, hence facilitating rapid model training and real-time inference.
- **RAM**: 64GB DDR4 memory provides sufficient capacity for handling large datasets and many model instances throughout the training, validation, and testing phases.
- **Operating System**: Ubuntu 20.04 LTS, the robust and widely used Linux version, provides seamless interoperability with prominent deep learning frameworks such as TensorFlow, PyTorch, and Keras.

This configuration was used to accelerate training cycles and eliminate computing limitations. It also facilitates the execution of computationally intensive activities such as batch loading high-resolution MRI images into memory and finetuning deep convolutional networks. The GPU proved crucial in expediting gradient updates and backpropagation, hence substantially reducing training time.

G. ACCURACY OVER EPOCHS

An essential metric of a model's ability to accurately classify data is accuracy. Figure 3 illustrates the model accuracy of ResNet50, InceptionV3, and VGG16 during the training process. All models exhibited a positive learning trajectory; however, the performance trends varied significantly across architectures.

By the end of the 10th epoch, ResNet50 achieved a peak training accuracy of **98%**, outperforming InceptionV3 (**96%**) and VGG16 (**95%**). This superior accuracy is directly attributed to the architectural advantage of ResNet50, which incorporates residual connections. These connections allow the model to learn identity mappings, thus minimizing the vanishing gradient issue that commonly restricts deep networks.

Residual learning allows for deeper architectures without degradation in performance, facilitating the capture of fine-grained features critical for distinguishing between tumor types. In contrast, while InceptionV3 utilizes inception modules to extract multi-scale features and VGG16 uses a consistent convolutional structure, both fall short in comparison to ResNet50's feature propagation capacity and representational depth.

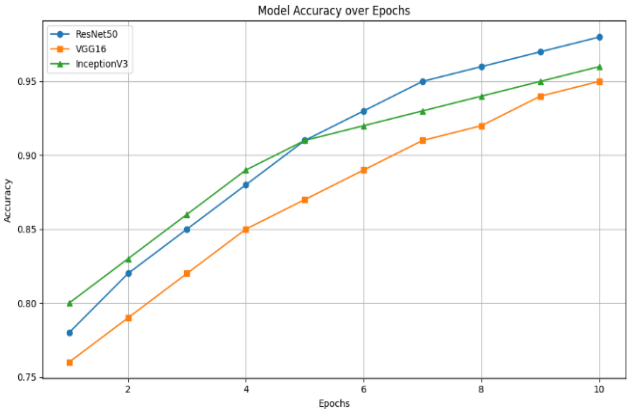

FIGURE 3: Accuracy Comparison.

H. LOSS OVER EPOCHS

Training loss, derived from the binary cross-entropy function, quantifies the discrepancy between predicted and actual labels. A declining loss curve signifies successful learning. As depicted in Figure 4, all models demonstrate a steady decrease in loss over epochs, but ResNet50 converged significantly faster and to a lower value.

Initially, all models began with a high loss (0.6), but ResNet50's loss sharply declined to **0.10** by the final epoch. In comparison, InceptionV3 and VGG16 plateaued at higher values of **0.12** and **0.15**, respectively. This rapid convergence in ResNet50 can be attributed to its advanced learning capacity, which stems from both depth and residual connections that facilitate effective feature reuse and error signal propagation.

The lower loss indicates better model confidence and generalization, reducing the likelihood of overfitting or underfitting—a crucial factor in medical imaging, where data diversity and feature subtlety are pronounced.

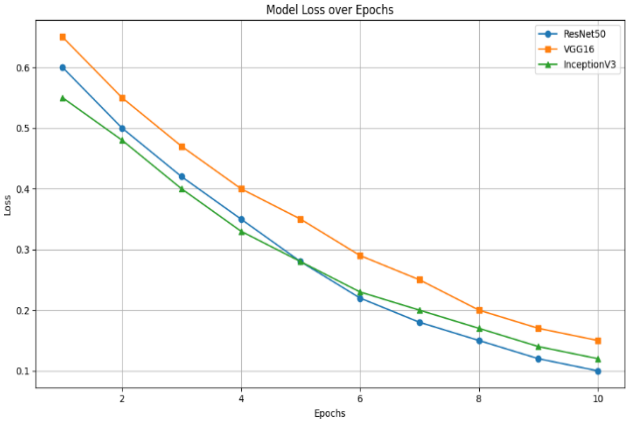

FIGURE 4: Loss over Epochs.

I. CONFUSION MATRIX (RESNET50)

The confusion matrix (Figure 5) presents a thorough analysis of classification results by comparing anticipated labels versus genuine class labels. For multi-class issues such as brain tumor classification, it is a vital diagnostic tool to assess class-specific model performance.

ResNet50's confusion matrix reveals a high number of **True Positives (TP) and True Negatives (TN)** across all four categories: glioma tumor, meningioma tumor, pituitary tumor, and no tumor. The little misclassification rate highlights its significant **sensitivity** (capacity to identify actual tumor cases) and **specificity** (ability to accurately differentiate non-tumor instances).

In therapeutic settings, where false negatives may delay treatment and false positives might lead to unnecessary interventions, this high level of accuracy is very crucial. The reliability of ResNet50's predictions indicates its potential as a trustworthy decision-support instrument in radiological diagnostics.

J. ROC CURVE

The Receiver Operating Characteristic (ROC) curve (Figure 6) illustrates the true positive rate in relation to the false positive rate across various categorization levels. The classification efficacy of a model is visually represented as a curve.

ResNet50 demonstrated exceptional discriminative ability between tumor and non-tumor occurrence with an **AUC (Area Under Curve)** score of **0.96**. The ROC curve demonstrates robust class separation, remaining far above the diagonal baseline despite data imbalance.

In high-stakes medical applications, such high AUC values confirm the model's capacity to distinguish subtle variations in MRI scans, which might be imperceptible to the human eye, thereby enhancing diagnostic accuracy.

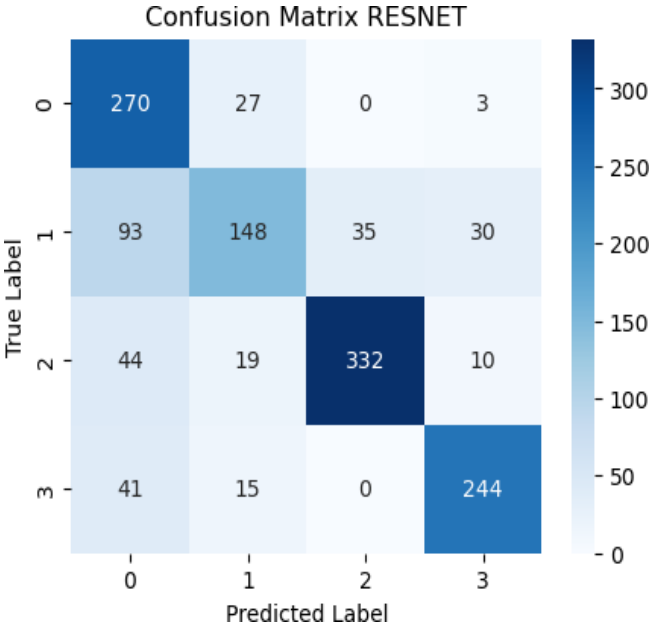

FIGURE 5: Confusion Matrix.

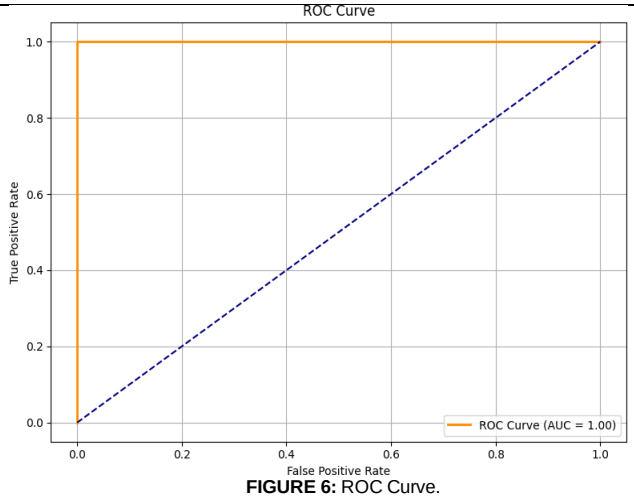

FIGURE 6: ROC Curve.

K. PRECISION-RECALL CURVE

The precision-recall curve (Figure 7) gives insights into the trade-off between precision (positive predictive value) and recall (sensitivity). It is particularly informative in cases of class imbalance, which is common in medical datasets.

ResNet50 maintained a consistently high balance between precision and recall throughout the range of thresholds. The large area under the curve (AUC) indicates that the model sustains high precision without compromising recall. This is crucial in a medical context, as high recall ensures tumor cases are not overlooked, while high precision minimizes the rate of false alarms.

Such robustness makes ResNet50 well-suited for deployment in environments where the consequences of diagnostic errors are significant, such as oncology departments and neurological clinics.

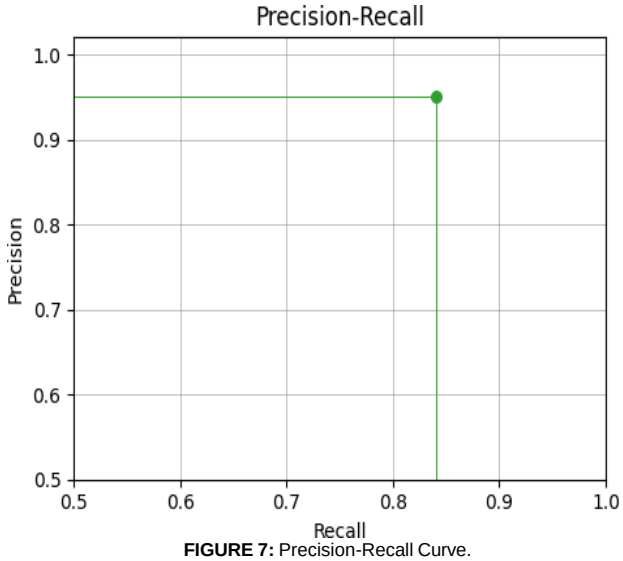

FIGURE 7: Precision-Recall Curve.

L. F1 SCORE COMPARISON

The F1 score, being the harmonic mean of recall and accuracy, provides a singular metric for assessing a model's sensitivity to the balance between these two measures. ResNet50 achieved the highest F1 score of **0.92**, as seen in Figure 8, followed by InceptionV3 at

0.90 and VGG16 at **0.88**.

This result underscores ResNet50's efficacy in addressing complex multi-class classification challenges, particularly when several tumor types exhibit overlapping visual traits. The equilibrium of its F1 score indicates that the model does not disproportionately favor one class over another, an essential attribute for fairness and equity in medical artificial intelligence applications.

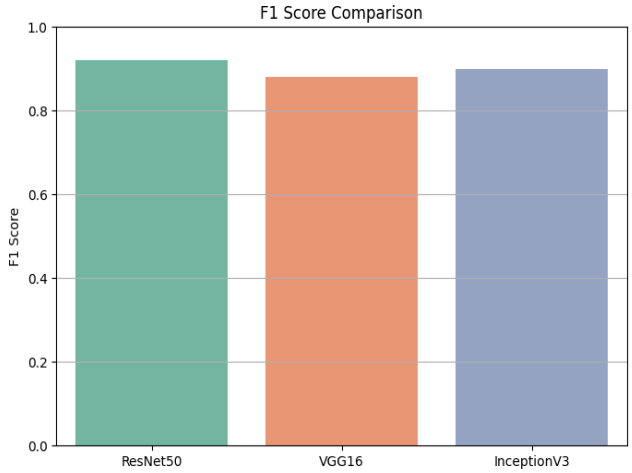

FIGURE 8: F1 Score Comparison.

M. AUC SCORE COMPARISON

Examining the bar graph (Figure 9) that compares AUC ratings further emphasized the efficacy of each model. ResNet50 achieved an AUC of **0.96**, leading the results, followed by InceptionV3 at **0.94** and VGG16 at **0.93**.

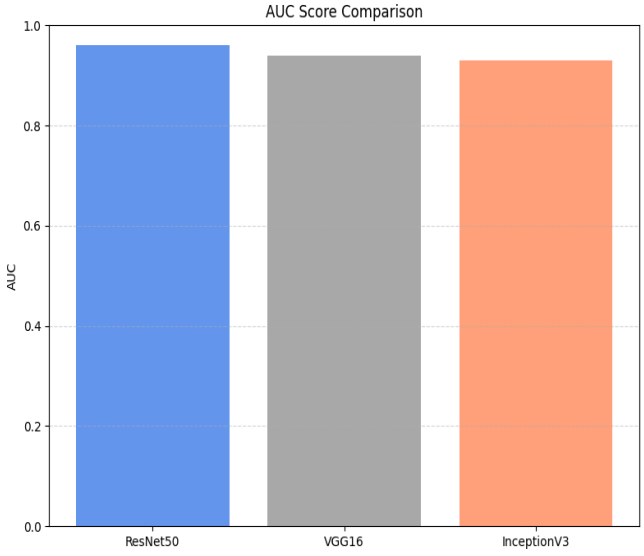

FIGURE 9: AUC Score Comparison.

These results validate ResNet50's reliable performance across many evaluation metrics and provide substantial evidence of its suitability for clinical applications. Its durability and flexibility are shown by its elevated AUC, rapid convergence, low misclassification rate, and robust precision-recall trade-off.

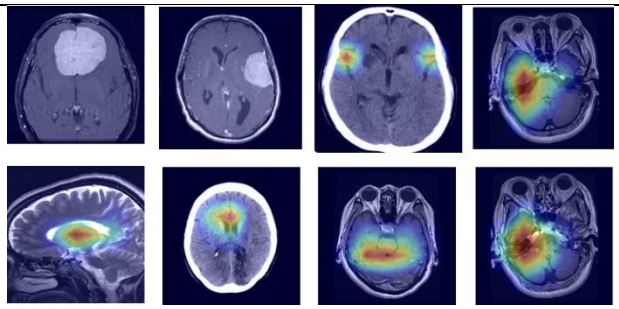

FIGURE 10: Tumor Localization via Grad-CAM.

We used the Grad-CAM (Gradient-weighted Class Activation Mapping) approach to elucidate the decision-making process of the ResNet18 classification model. Grad-CAM superimposes a heatmap over the original brain MRI, emphasizing the regions most likely to influence the model's predictions.

A pre-trained ResNet18 model was used, which was modified to perform inference on brain MRI images. The model was set to evaluation mode, and Grad-CAM visualizations were generated for each input image. Specifically, gradients were extracted from the last convolutional layer (i.e., 'layer4.1.conv2') relative to the predicted class. These gradients were pooled and weighted against the corresponding feature maps to produce a class-discriminative localization map.

The resulting heatmaps were resized and superimposed on the original MRI scans, revealing regions of attention. As shown in Figure 10, the model focuses primarily on hyperintense regions commonly associated with gliomas, meningiomas, or pituitary tumors. In correctly classified cases, the attention maps align with tumor regions marked by radiologists, validating both the performance and interpretability of the deep model.

V. FINDINGS

The comparative examination of three state-of-the-art deep learning models, ResNet50, InceptionV3, and VGG-16, revealed crucial insights about their potential for brain tumor classification using Magnetic Resonance Imaging (MRI) information. Each model was rated based on numerous performance measures that include accuracy, training loss, F1score, Area Under the Curve (AUC), confusion matrix analysis, precision-recall trade-off, and convergence speed.

A. OVERALL MODEL PERFORMANCE

Among all the models evaluated, **ResNet50 consistently emerged as the superior architecture**. By the 10th epoch, it achieved a peak classification accuracy of **98%**, surpassing InceptionV3 at **96%** and VGG16 at **95%**. The residual learning architecture of ResNet50 accounts for its remarkable accuracy by facilitating deeper feature learning while mitigating the risk of vanishing gradients. Among the four tumor types, glioma, meningioma, pituitary tumor, and absence of tumor, its ability to extract complex and distinctive features was essential for their differentiation.

B. TRAINING EFFICIENCY AND CONVERGENCE BEHAVIOR

ResNet50 exhibited the most rapid convergence during training, therefore reducing the binary cross-entropy loss from **0.6 to 0.10** compared to **0.12** for InceptionV3 and from **0.15** relative to VGG16. ResNet50 is suitable for time-sensitive clinical environments where rapid model training and retraining are essential, since its fast convergence demonstrates excellent learning dynamics. The model exhibited few signs of overfitting and remained stable across epochs.

C. PRECISION, RECALL, AND F1 SCORE

ResNet50 achieved the highest F1 score of **0.92** for classification quality, indicating an effective equilibrium between recall and accuracy. InceptionV3 and VGG16 achieved F1 scores of **0.90** and **0.88**, respectively. Our results validate ResNet50's robustness in addressing class imbalance and atypical tumor classes, which is particularly relevant in real-world datasets where these issues are prevalent.

D. DISCRIMINATORY CAPABILITY AND ROC-AUC ANALYSIS

The **AUC value of 0.96** for ResNet50 clearly demonstrates its discriminative capability. This statistic illustrates the model's efficacy in distinguishing classes at certain threshold levels. The reliability of ResNet50 in clinical decision-making contexts, particularly when false positives or false negatives might have serious repercussions, was substantiated by its ROC curve, which consistently remained above the diagonal baseline.

E. CONFUSION MATRIX INTERPRETATION

The confusion matrix of ResNet50 demonstrated commendable sensitivity (true positive rate) and specificity (true negative rate), indicating minimal misclassifications across all four classes. This exceptional diagnostic capability indicates the model's suitability for incorporation into a Computer-Aided Diagnosis (CAD) system, therefore assisting radiologists in accurately identifying brain tumors with little error.

F. PRECISION-RECALL TRADE-OFF

ResNet50 exhibited a robust trade-off curve in the precision-recall analysis, indicating its ability to preserve accuracy while maintaining recall. In medical imaging, strong recall ensures the identification of almost all tumor cases, while high accuracy minimizes unnecessary false alarms that might lead to unwarranted therapeutic interventions, making this aspect very important.

G. MODEL EFFICIENCY AND PRACTICAL APPLICABILITY

Despite all three models using pre-trained CNNs and transfer learning, ResNet50 yielded a compelling combination of efficiency and performance. Despite being a more complex network, fine-tuning techniques contributed to a reduction in computing expenses. Its minimal error rates and high accuracy, coupled with rapid training durations, make it an excellent option for deployment in real-time, resource-constrained clinical settings.

Table II presents the comparative outcomes across all primary performance metrics.

The findings of this study indicate that ResNet50 is the most compelling design for multi-class brain tumor classification based on MRI. Its effectiveness across all metrics designates it as a reliable and efficacious approach for clinical implementation. Its potential as a foundational model in forthcoming AI-assisted diagnostic systems is underscored by its resistance to overfitting, equitable classification across categories, and suitability for resource-constrained settings.

TABLE 2: Performance Comparison of Deep Learning Models for Brain Tumor Detection

Metric	ResNet50	InceptionV3	VGG16
Final Accuracy (%)	98	96	95
Final Loss	0.10	0.12	0.15
F1 Score	0.92	0.90	0.88
AUC Score	0.96	0.94	0.93
Precision-Recall	High	Moderate-High	Moderate
Convergence Speed	Fastest	Moderate	Slower
Confusion Matrix Result	Excellent (few errors)	Good	Good

VI. CONCLUSION

This study conducted a comparative examination of three deep learning models, ResNet50, VGG16, and InceptionV3, with transfer learning utilizing MRI data for brain tumor detection and classification, also used Grad-CAM the explainable artificial intelligence (XAI) strategy, to boost model transparency and interpretability. ResNet50 has much superior accuracy, F1-score, AUC, and convergence rate compared to the alternatives. This may be attributed to its residual connections, which provide more efficient gradient propagation and deeper representation learning, both crucial in medical image processing, where minor differences are significant.

Future research will aim to enhance the model's generalizability across multi-center datasets with varying image collection protocols. Additionally, enhancing interpretability for physicians might include the use of explainable artificial intelligence (XAI) systems such as Integrated Gradiant (IG), DeepLIFT, and Score-CAM. Moreover, the model may be further extended for multi-class classification, including several tumor grades or the segmentation of tumor regions and the size of the tumor. An alternative approach to facilitate system deployment in remote and resource-constrained environments is the integration with mobile platforms and real-time cloud-based inference engines.

DATA AVAILABILITY

The datasets included in this study is combination of SARTAJ, Figshare, Br35h, and publicly accessible and widely employed in brain tumor detection [1] studies. The

dataset is open-source and is used in compliance with its respective data usage policies.

[1]https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

REFERENCES

[1] I. Jahan and M. L. Rahman, "Detection of brain tumor using Internet of things," 2018.

[2] R. Zhang, H. Luo, W. Chen, and Y. Bai, "Review of deep learning-driven MRI brain tumor detection and segmentation methods," Advances in Computer, Signals and Systems, 2023.

[3] C.-C. Peng and B.-H. Liao, "Classify brain tumors from mri images: Deep learning-based approach," 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS), pp. 5–8, 2023.

[4] P. P. Kalluri, V. M. Lomte, A. Kangude, P. Kharate, and K. Tibe, "Brain tumor detection and segmentation using bit-plane and unet," 2021.

[5] M. N. Putri, I. Katili, A. Hariri, T. A. Budiarti, and G. M. Wibowo, "Perbandingan pegukuran volume tumor brain mri menggunakan teknik manual dan metode active contour," Jurnal Imejing Diagnostik (JImeD), 2021.

[6] J. Y. Tan, J. Y. Thong, Y. H. Yeo, K. Mbenga, and S. Saleh, "Gender, racial, and geographical disparities in malignant brain tumor mortality in the United States," Oncology, 2024.

[7] S. Das, M. Sarder, S. Das, D. H. Tanvir, S. T. Aziz, and A. Islam, "Deep learning-assisted MRI image segmentation and classification for precise brain tumor analysis," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), pp. 1–6, 2023.

[8] K. Thiruvenkadam, V. Ravindran, and A. Thiyagarajan, "Deep learning with xai based multi-modal mri brain tumor image analysis using image fusion techniques," 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies, pp. 1–5, 2024.

[9] R. Ramachandran, A. Prasad, R. R. Prasad, V. S. Nair, and L. G. Pillai, "Medical image analysis using distributed deep learning models," 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), pp. 1–5, 2023.

[10] K. J. Johnson, L. Bauchet, S. S. Francis, J. A. Hainfellner, C. Kruchko, C. C. Lau, Q. T. Ostrom, M. E. Scheurer, and Y. Yuan, "Pediatric brain tumors: Origins, epidemiology, and classification the 2022 brain tumor epidemiology consortium meeting report.," Clinical neuropathology, 2023.

[11] R. Mohan, J. Wahyuhadi, and N. W. Tirthaningsih, "The profile of brain tumor cases in rsud dr soetomo, surabaya," 2021.

[12] A. Gulhane and A. Velmurugan, "Design and analysis for advancements in brain tumor detection model by using machine learning techniques," 2024 8th International Conference on Inventive Systems and Control (ICISC), pp. 13–18, 2024.

[13] S. Jain and V. Jain, "Machine learning and deep learning methods in brain tumor classification: A decade: Systematic literature review," Intelligent Data Analysis, 2024.

[14] A. Pimpalkar, P. Tembhurne, A. Ingle, V. Gosawi, and P. Patle, "Brain tumor detection and classification using machine learning: A comprehensive survey," International Research Journal of Modernization in Engineering Technology and Science, 2023.

[15] P. K. Kushwaha, A. Rajput, S. Aggrawal, S. P. Dwivedi, A. Srivastava, and S. Singh, "Brain tumour detection using machine learning," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), vol. 6, pp. 557–560, 2023.

[16] M. Selvi, K. Gokul, and D. Dhivin, "Classification of brain tumor detection techniques a review," 2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1525– 1531, 2024.

[17] M. O. Arowolo, W. O. Ajayi, P. O. Olawoye, O. Babajide, H. E. Aigbogun, M. D. Salawu, M. O. Adebiyi, and A. A. Adebiyi, "Empowering healthcare with ai: Brain tumor detection using mri and multiple algorithms," 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), pp. 1–11, 2024.

[18] Q. Qusay, Q. W. Bsoul, F. yassin Salem Al jawazneh, R. W. Bsoul, D. S. Abdelminaam, M. A. Abd-Elghany, Y. Alkady, and I. A. E. Gomaa, "Brain tumor detection: Integrating machine learning and deep learning for robust brain tumor classification," Journal of Intelligent Systems and Internet of Things, 2025.

[19] J. J, S. Raghul, N. A. Kumar, K. R. Desai, V. Chourasia, and A. K. Agrawal, "Deep learning-enhanced mri for brain tumor detection and characterization," 2023 9th International Conference on Smart Structures and Systems (ICSSS), pp. 1–6, 2023.

[20] V. L. Castelino, M. Vishnu, K. Shetty, P. Jain, V. Kamath, and V. Thanthri, "Brain tumor detection using machine learning," 2024 Second International Conference on Data Science and Information System (ICDSIS), pp. 1–8, 2024.

[21] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345–1359, 2010.

[22] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" NeurIPS, vol. 27, pp. 3320–3328, 2014.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[25] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2818–2826, 2016.

Date available online: 09-01-2026
Vol. 3, Issue 1 (July – December 2025)

DOI: https://doi.org/10.24312/ucp-jeit.03.02.708

End-to-End Motorcycle Violation Detection with Region-Specific Automatic License Plate Recognition

Mohamed Rafi Atheek[1], Mohamed Buhary Fathima Anizul Fathool[2], Atif Ishaq Khan[3]
[1]Department of Computer Science, GC University, Lahore, Pakistan (e-mail: mhdatheek136@gmail.com)
[2]Department of Computer Science, GC University, Lahore, Pakistan (e-mail: anizulfathool@gmail.com)
[3]Department of Computer Science, GC University, Lahore, Pakistan (e-mail: atif.ishaq@gcu.edu.pk)

Corresponding author: Atif Ishaq Khan (e-mail: atif.ishaq@gcu.edu.pk).

ABSTRACT

We present a region-aware, end-to-end motorcycle violation detection pipeline tailored to traffic conditions in Punjab, Pakistan, which integrates three YOLOv11-based components into a unified framework: motorcycle violation detection (MCVD) for helmet compliance and multi-rider analysis, license plate detection (LPD), and license plate character detection (LPCD). The system integrates lightweight object detection, BoT-SORT-based tracking, and character-level recognition, supported by a synthetic-toreal adaptation strategy that combines large-scale synthetic data with limited real samples. Two specific datasets are published, a 40,000-sample synthetic Punjab license plate dataset (PS-LPCD) and a 650-sample real-world dataset (PR-LPCD), which are publicly released in order to encourage research development and adaptation to the region. Class consolidation enhanced MCVD performance (weighted average F1 score: 0.77) and the LPD model performed at $mAP_{50} = 0.99$. Two-stage fine-tuning on synthetic and real samples allowed LPCD to reach a character accuracy of $\approx$ 98% and a full-plate recognition rate of $\approx$ 90.7%, both surpassing EasyOCR and PaddleOCR, while also achieving lower per-plate latency. With a single motorcycle per frame, the sequential pipeline maintains a throughput of $\approx$ 9.5 FPS; the throughput reduces in scenes where there are many motorcycles. These findings indicate that synthetic pretraining, together with a small real fine-tuning, can be used to obtain a powerful, scalable, and region aware automatic license plate recognition (ALPR) system, which provides a reproducible method for detecting traffic violations across a variety of license-plate formats.

INDEX TERMS ALPR, Helmet compliance, License plate recognition, Motorcycle violation detection, Multi-rider counting, Punjab, Synthetic dataset, YOLOv11

I. INTRODUCTION

Motorcycle-related traffic violations are a major contributor to road injuries and fatalities worldwide. The World Health Organization (WHO) reports that motorcyclists account for 21% of all road traffic deaths [1]. In Pakistan, motorcycles are a dominant mode of transport and are disproportionately represented in crash statistics [2], [3], underscoring the need for effective, region-aware monitoring and enforcement systems.

Although the previous literature has already generated precise approaches to individual tasks such as helmet detection [4], multi-rider counting [4], and automatic license plate recognition (ALPR) [5], most systems address these sub-tasks in isolation. Moreover, reliance on generic datasets and off-the-shelf OCR engines (e.g., EasyOCR, Tesseract) limits robustness in regions where license plate formats and scripts vary. This constrains both applicability and reproducibility in real-world deployments.

To fill these gaps, we propose an end-to-end motorcycle violation detection pipeline based on YOLOv11. We emphasize that this work does not introduce new detection architectures or learning algorithms. Instead, the novelty lies in a systems-level integration of existing state-of-the-art components, combined with region-specific dataset design, synthetic-to-real adaptation, and deployment-oriented evaluation. The system integrates three modules: MCVD (Motorcycle Violation Detection) for helmet-use and multi-rider detection, LPD (License Plate Detection) for plate localization, and LPCD (License Plate Character Detection) for character recognition. Together, these components form a practical, reproducible, and region-aware motorcycle violation detection pipeline.

OUR MAIN CONTRIBUTIONS ARE AS FOLLOWS:

- A unified, end-to-end YOLOv11-based pipeline integrating helmet detection, rider counting, and ALPR through system-level design.
- Two new Punjab-specific license plate character detection datasets (synthetic and real) released to support region-aware ALPR research and reproducibility.
- A lightweight character-level detection approach that improves ALPR robustness compared to off-the-shelf OCR engines.
- Models trained and evaluated on augmented public

datasets to ensure both reproducibility and regional applicability.

II. RELATED WORK

Initially, the detection of helmet-use was based on handcrafted features and classical classifiers. As an example, SVMs were used on the histograms of head regions with background subtraction and projection profiling [6], moving blob extraction with K-Nearest Neighbor classification [7], and combined LBP, HOG, and Hough descriptors, achieving an accuracy of 94.23% [8]. These methods were generally sensitive to lighting, occlusion, and crowding.

Deep learning has enhanced its strength and enabled joint tasks. CNN-based classifiers, for example, achieved high accuracy (96.6%) and F_1-score (94.6%) [9]. Various other pipelines, including YOLO-based and alternative approaches, have also been applied for helmet detection and multi-rider identification [4], [10]–[15]. Nevertheless, although the above-mentioned methods work well in their intended applications, they are typically not integrated with the ALPR systems, which restricts their use in end-to-end motorcycle violation detection pipelines.

In the case of Pakistan, the research on the topic has focused on individual tasks and not on comprehensive end-to-end motorcycle violation detection. Deep learning models have been successful in identifying the location of the helmet on a surveillance video with high accuracy [16], [17], while ALPR systems have focused on license plate localization, character segmentation, and OCR [18], [19]. Motorcycle-based end-to-end pipelines involving the detection of helmet violations, multi-rider, and region-specific license plate recognition are still uncommon. This gap is addressed in our work, where a unified framework is proposed, which is specific to Punjab, Pakistan.

Several works outside Pakistan have integrated helmet detection with ALPR in end-to-end pipelines. In some cases, evaluation relied on proprietary datasets or generic OCR systems, and regional plate variations were not always addressed [20]–[24].

Synthetic data has emerged as a viable solution to the scarcity and privacy issues of license plate datasets. Template-based methods [25], rendering pipelines [26], and diffusion models [27] have shown measurable gains in recognition accuracy. Based on such methods, we generate a template-based synthetic dataset of character-level annotated license plates specifically for the Punjab, Pakistan region, complemented with manually labeled real images for evaluation. These findings, along with benchmarking studies [28], demonstrate the viability of synthetic plate generation as a reliable supplement to real data.

III. METHODOLOGY

A. SYSTEM OVERVIEW

The suggested end to end Motorcycle Violation Detection system will be used to monitor helmet compliance, multi riders, and license plate recognition. The pipeline will be composed of motorcycle detection, tracking, license plate detection, character recognition, and violation classification as illustrated in Figure 1.

The detection of motorcycles was conducted with the help of the YOLOv11 model that was trained on the COCO dataset [29]. To optimize efficiency, the system first employs lightweight detection and BoT-SORT tracking [30] to identify candidate violation frames where motorcycles and license plates are both visible and potentially readable. Only these frames are then processed with the heavier MCVD YOLOv11m model, ensuring a balance between accuracy and computational cost. This design prevents violations from being logged on unreadable plates, which is essential for reliable automated enforcement.

B. DATASETS AND PREPROCESSING

1) Motorcycle Violation Dataset (HELMET)

The HELMET dataset [4] is a widely used benchmark for helmet-use and multi-rider detection and was adopted for training the MCVD model. It comprises 91,000 annotated frames with 283,377 labeled object instances spanning 36 fine-grained classes. As is typical of real-world traffic data, this detailed class structure introduces a substantial class imbalance, with several safety-critical violations occurring far less frequently than compliant riding behaviors.

To address this imbalance, two complementary strategies were employed. First, class consolidation was performed to simplify the label space and better reflect traffic enforcement practices. The front-child passenger (P0) class was removed, and all cases involving three or more riders were grouped into a single *MoreThanTwoRider* category, since any rider count exceeding two constitutes a violation regardless of helmet usage. Second, a targeted sampling approach was applied during training, where horizontal flipping augmentation was restricted to underrepresented classes (those with fewer than 20,000 samples). This selectively increases the representation of minority classes without distorting the natural distribution of dominant categories.

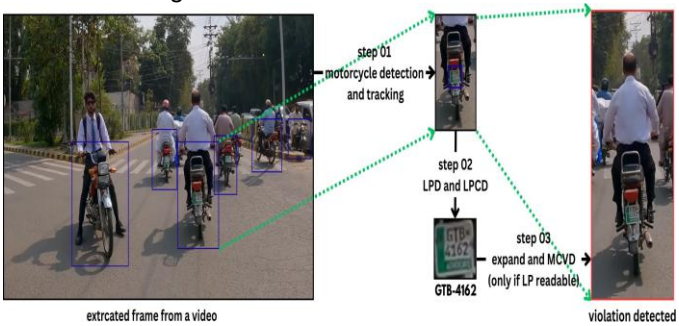

FIGURE 1. Pipeline of the proposed End-to-End Motorcycle Violation Detection system, including motorcycle detection, tracking, license plate recognition, and violation classification.

Following consolidation and augmentation, the dataset contained 318,131 annotated instances across seven violation-relevant classes. The class distributions before and after consolidation are provided in Appendix A.

The dataset was divided into 70% for training, 10% for validation, and 20% for testing. Since the original dataset consists primarily of CCTV footage in which motorcycles often appear at a distance, near-camera close-up views of riders and helmets are underrepresented. To address this limitation and improve robustness to real-world scale variations, we applied a two-step augmentation strategy:

1) **Close-up cropping:** Every 10th frame containing a motorcycle was cropped with a padding of 0.1 (as a fraction of the bounding-box size) to synthetically generate near-camera views while preserving annotation coordinates.
2) **Super-resolution enhancement:** The resulting low-resolution close-up crops were enhanced using RealESRGAN [31] to recover fine-grained details and improve object detectability.

In addition, mosaic blending was applied to simulate dense traffic conditions and improve robustness to occlusion and scale variation.

2) License Plate Detection Dataset (UFPR-ALPR)

In the case of license plate detection, we resorted to the UFPR-ALPR dataset that consists of 4,500 annotated images that represent various types of vehicles [5]. Each annotation included a license-plate bounding box and metadata (vehicle type, camera type, lighting conditions). We focused on plates that are visible on motorcycles and on vehicle types that are relevant for LPD.

To improve regional relevance for Punjab, Pakistan, we generated synthetic Punjab-style license plates and replaced the original plates in the images while preserving plate aspect ratios (single-line vs. double-line), as shown in Figure 3. This augmentation doubled the dataset to 9,000 images. Following the original dataset recommendations, the split was 40% training (3,600 images), 40% validation (3,600 images), and 20% test (1,800 images) [5].

Training augmentations for the LPD model included mosaic blending, shear, perspective deformation, and limited horizontal flipping. This set of augmentations simulates viewpoint variation and minor geometric distortions while preserving plate legibility.

3) License Plate Character Dataset (PS-LPCD and PR-LPCD)

To train a robust character-level detector, we created the Punjab Synthetic License Plate Character Dataset (PS-LPCD) and a complementary real-world dataset, the Punjab Real License Plate Character Dataset (PR-LPCD). PS-LPCD contains 40,000 synthetic images generated across four Punjab plate templates, while PR-LPCD comprises 650 annotated crops extracted from the PK-Number-Plates-V3 collection [32]. After filtering for Punjab templates, 500 samples were reserved for fine-tuning and 150 for final testing. PSLPCD was split into 80% for training and 20% for validation. Sample synthetic examples are shown in Figure 4.

Both datasets are freely available for research purposes as part of the Punjab Pakistan Synthetic and Real License Plate Character Datasets (P-LPCD), available at Zenodo (https://doi.org/10.5281/zenodo.17182320) [33].

PS-LPCD contains 40,000 synthetic images equally divided among four Punjab plate templates (front/back × old/new; 10,000 images per template). We annotated 37 classes: digits 0–9, uppercase letters A–Z, and a special class “PUNJAB” used to detect decorative or regional markers and to filter irrelevant glyphs. Synthetic images were randomized with the following augmentations to emulate real capture artifacts:

- Spatial transforms applied with probability 0.7: translation ±10 pixels, shear ±15°, rotation ±15°.
- Perspective warp (small magnitude) to simulate viewpoint changes.
- Photometric and environmental noise: dirt, dust, Gaussian noise, and blur.
- Motion blur applied with probability 0.5; kernel size n chosen randomly from odd integers in [1, 29]. The discrete motion-blur kernel $K_{i,j}$ is defined as

$$K_{i,j} = \begin{cases} \frac{1}{n}, & \text{if } i = \lceil n/2 \rceil \quad \text{(horizontal blur row)}, \\ \frac{1}{n}, & \text{if } j = \lceil n/2 \rceil \quad \text{(vertical blur column)}, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

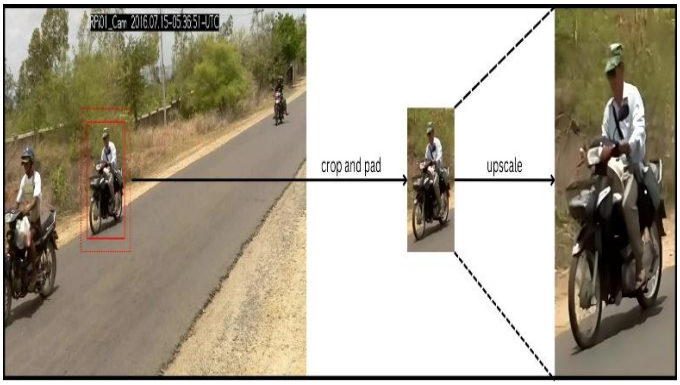

FIGURE 2. Two-stage super-resolution augmentation process. The first stage generates synthetic close-up views through cropping, and the second stage applies Real-ESRGAN to enhance visual details for improved detection performance.

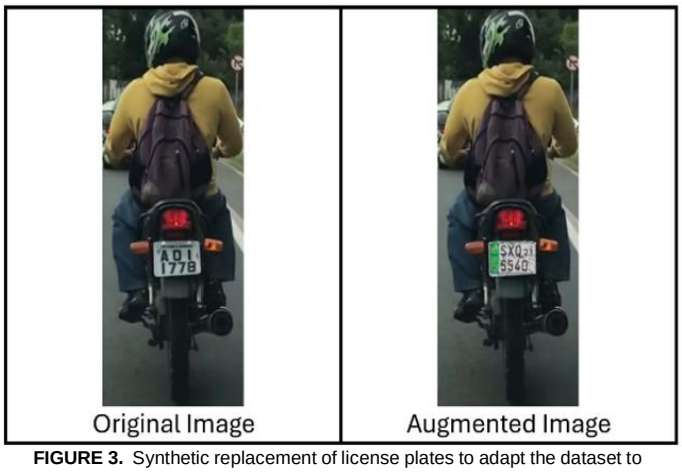

FIGURE 3. Synthetic replacement of license plates to adapt the dataset to regional characteristics.

This kernel produces a uniform linear blur across the central row (horizontal) or column (vertical), modeling motion along the principal axes.

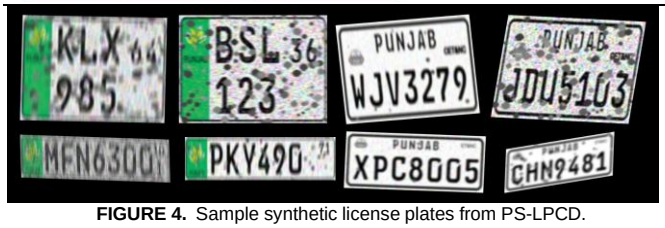

FIGURE 4. Sample synthetic license plates from PS-LPCD.

C. MODEL ARCHITECTURE AND TRAINING

The suggested system utilizes the models based on the YOLOv11 to perform three main tasks, namely, Motorcycle Violation Detection (MCVD), License Plate Detection (LPD), and License Plate Character Detection (LPCD). Two model versions were used: YOLOv11n, a small model with real time inference, and YOLOv11m, a large model that is designed to achieve the highest accuracy possible at the cost of computational performance [34].

All models were initialized with weights pretrained on the COCO dataset [35], a large-scale benchmark dataset. Pretraining provides transferable and generalized feature representations that improve performance across various vision tasks [36]. Task-specific data preprocessing and augmentation strategies were then applied to enhance robustness under challenging traffic conditions, including occlusion, motion blur, and varied viewpoints. For example, MCVD training incorporated moderate mosaic blending to improve detection in dense scenes, while LPCD training disabled mosaic augmentation and horizontal flips to preserve character orientation. The LPD training used a mixed augmentation, which consisted of mosaic blending, shear, perspective deformation, and restricted flips to preserve the geometry of the license plates. In addition to these custom settings, all models utilized the default augmentation pipeline provided in the YOLOv11 documentation [29].

The training was done on a 64-bit system with an NVIDIA RTX 2060 GPU (VRAM: 6GB) and an AMD Ryzen 7 4800H CPU (with 8 cores and 16 threads), along with 40 GB of RAM, and operated under the Windows operating system. Hyperparameters were tuned to balance accuracy and inference efficiency. In particular, the MCVD model was trained using the SGD optimizer with momentum, as the HELMET dataset is large and SGD is known to offer better generalization on large datasets. In contrast, the LPD and LPCD models were trained using the Adam optimizer, since their datasets are medium or small in size, where adaptive methods such as Adam converge faster [37]. The overall training configuration is summarized in Table 1.

TABLE 1. Training configurations for MCVD, LPD, and LPCD models

Model	Epochs	Batch Size	Optimizer
MCVD (YOLOv11n)	20	16	SGD
MCVD (YOLOv11m)	15	8	SGD
LPD (YOLOv11n)	50	16	Adam
LPD (YOLOv11m)	20	8	Adam
LPCD (YOLOv11n)	20	16	Adam

All models were trained with an image size of 640 × 640.

Complete training and validation curves, including loss (box, cls, dfl), mAP, precision, and recall for all YOLOv11 models, are provided in Appendix B for reference and reproducibility.

In the case of LPCD, domain adaptation was performed through a two-step fine-tuning process on the Punjab Real License Plate Character Dataset (PR-LPCD) following the initial training, as shown in Table 1. In Stage One, the first eight layers were frozen, and the model was trained for 10 epochs with a learning rate of 1×10^{-5} and a batch size of 8. In Stage Two, only the first four layers remained frozen, and training continued for another 10 epochs with a reduced learning rate of 1×10^{-6}. This gradual unfreezing approach allowed the model to fit well to real-world data and alleviate overfitting while retaining the generalizable characteristics acquired in synthetic training.

D. LICENSE PLATE CLASSIFICATION

The LPCD pipeline includes a license plate layout classification step in order to allow the correct sequencing of the detected characters, separating single-line and double-line plates. This difference is essential, as the position of the characters varies dramatically across layouts.

The classification is based on a normalized vertical variation measure that is calculated using the bounding boxes of the identified characters. Let y_i denote the vertical center of the i^{th} character, and h_i its bounding box height. The metric is defined as:

$$\text{Normalized Variation} = \frac{\sigma_y}{\mu_h} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}}{\frac{1}{N}\sum_{i=1}^{N} h_i} \quad (2)$$

where σ_y represents the standard deviation of the vertical centers, μ_h the mean character height, and N the total number of detected characters.

Plates where the normalized variation was more than 0.45 were defined as double-line because the characters were more vertically spread. Values below this threshold indicated single-line plates. The threshold was chosen empirically in steps of 0.05 by comparing the accuracy of classification on the PR-LPCD data.

After classification, single-line plates were read sequentially, whereas double-line plates were parsed line by line. For double-line plates, a vertical midline, computed as the average of all vertical points, separates the two lines, which are read independently. Additional heuristics were incorporated to improve robustness, such as identifying the smaller-sized final two digits of the registration year that commonly appear on Punjab single-line plates.

Figure 5 illustrates the process, showing bounding box distributions for singleand double-line plates. A vertical yellow midline separates the two lines in double-line plates, while red dots mark each character's center, and dotted red lines indicate the spread from the midline.

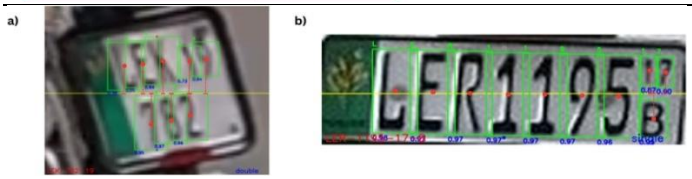

FIGURE 5. Normalized Vertical Variations for Single-Line and Double-Line License Plates. Red dots indicate character centers, dotted red lines show the vertical spread, and the yellow midline separates the two lines in double-line plates.

IV. RESULTS AND DISCUSSION

A. EVALUATION METRICS

The performance of the proposed MCVD system was evaluated using metrics that reflect both detection accuracy and computational efficiency. For object detection, the primary metric is mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), which quantifies the model's ability to correctly localize and classify objects. Formally, for C classes and AP_c representing the average precision for class c:

$$\text{mAP@0.5} = \frac{1}{C}\sum_{c=1}^{C} \text{AP}_c \tag{3}$$

To provide a stricter assessment of detection robustness, mAP averaged over IoU thresholds from 0.5 to 0.95 in 0.05 increments (mAP@[0.5:0.95]) is computed as:

$$\text{mAP@[0.5:0.95]} = \frac{1}{10}\sum_{t=0.5}^{0.95} \text{mAP@}t \tag{4}$$

Given the residual imbalance across violation categories, the weighted F1-score was used as the primary evaluation metric for MCVD to ensure fair performance assessment across both frequent and rare classes. This is defined as:

$$F1_{\text{weighted}} = \sum_{c=1}^{C} w_c \cdot F1_c, \quad w_c = \frac{n_c}{\sum_{i=1}^{C} n_i} \tag{5}$$

where n_c denotes the number of true samples in class c and $F1_c$ is the F1-score for that class.

In addition, standard evaluation metrics such as accuracy, precision, recall, and F1-score are computed to provide additional insights into the system's performance. These are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

where TP , FP , FN , and TN represent true positives, false positives, false negatives, and true negatives, respectively.

To assess computational efficiency, inference speed was measured in frames per second (FPS). Higher FPS values indicate faster processing, which is crucial for real-time traffic monitoring applications.

B. MOTORCYCLE VIOLATION DETECTION (MCVD) MODELS

Three YOLO-based MCVD models were trained to investigate the effects of model size and architecture and to identify the best-performing model. Their performance was also compared with the CNN-MTL baseline (CNN-based Multi-Task Learning for helmet detection) proposed by Lin *et al.* [4], as shown in Table 2.

Consolidating the original 36 classes into 7 core violation classes significantly improved detection performance. YOLOv11n trained on 7 classes achieved an mAP@50 of 0.6584, nearly doubling YOLOv8n's 0.3514 trained on all 36 classes. This reduction in classes helped reduce label noise and class imbalance, thereby boosting accuracy and model focus. Notably, even with the same 36 classes, YOLOv8n outperformed the CNN-MTL baseline (F1 score 0.70 vs. 0.673), likely due to the more efficient and advanced YOLO architecture. YOLOv11m further improved performance, reaching an mAP@50 of 0.7127 and a weighted F1 score of 0.77, demonstrating the combined benefits of class consolidation and targeted minority-class augmentation in mitigating class imbalance and improving detection robustness.

For further analysis, the consolidated classes were grouped into *Non-Violation* (DHelmet, DHelmetP1Helmet) and *Violation* (DNoHelmet, DNoHelmetP1NoHelmet, DHelmetP1NoHelmet, DNoHelmetP1-Helmet, MoreThanTwoRider) categories, yielding the weighted average results shown in Table 3.

These results further illustrate the effectiveness of the proposed imbalance mitigation strategies, as minority violation classes benefit from improved recall without sacrificing precision on dominant non-violation categories.

The models show higher precision, recall, and F1-scores for compliant rider classes, effectively reducing false positives and improving the classification of non-violators. Inference speed tests on 1,000 random test images indicate that YOLOv11m achieves near real-time performance at approximately 25 frames per second (FPS), comparable to YOLOv11n's ~27 FPS, maintaining efficiency despite increased complexity.

Despite the strong overall performance, several failure

TABLE 2. Performance comparison of MCVD models, including YOLO variants and CNN-MTL baseline

MCVD Model	Precision	Recall	mAP@50	mAP@[50–95]	Classes	Weighted Average F1 Score
CNN-MTL [4]	–	–	–	–	36	0.673
YOLOv8n	0.3803	0.3889	0.3514	0.3104	36	0.70
YOLOv11n	0.6505	0.6157	0.6584	0.5915	7	0.72
YOLOv11m	0.7096	0.6664	0.7127	0.6517	7	0.77

cases were observed in challenging real-world scenarios. In highly crowded scenes, riders positioned very close to each other can be confused, leading to false predictions due to inter-instance occlusion. When occlusion occurs with nonmotorcycle objects, the model generally remains robust and is able to correctly distinguish riders. However, for distant motorcycles, headwear such as caps or hats is occasionally misclassified as helmets due to limited spatial resolution and visual similarity. Representative failure cases are illustrated in Figure 6.

A B C

FIGURE 6. Sample failure cases of the MCVD model. Annotations A and C illustrate misclassifications in highly crowded scenes involving closely spaced motorcycles, while B shows confusion between a cap and a helmet for distant rider instances.

In summary, label consolidation, enhanced training methods, and tailored augmentations enable YOLOv11-based MCVD models to deliver superior accuracy, precision, and practical deployment readiness, outperforming earlier state-of-the-art approaches.

C. LICENSE PLATE DETECTION (LPD) MODELS

The proposed LPD YOLOv11 models were evaluated on the UFPR-ALPR dataset [5], with results summarized in Table 4.

The lightweight YOLOv11n, trained for 50 epochs, slightly outperformed YOLOv11m, which was trained for 20 epochs, likely due to the longer training duration. Both models achieved high precision, recall, F1 score, and mAP, demonstrating accurate and well-localized license plate detection. While Laroca *et al.* [5] achieved a marginally higher recall (98.33% vs. 97.89%), our models provide complete metric coverage and near-perfect localization (mAP@0.5 = 0.9910). The mAP averaged over IoU thresholds 0.5 to 0.95 (mAP@[0.5–0.95]) reached 0.8485 for YOLOv11n and 0.8295 for YOLOv11m, highlighting robust detection across varying localization criteria and benefiting subsequent LPCD tasks.

These results were obtained on an augmented UFPRALPR dataset, including synthetic regional plates (see Section III-B2), which enhanced dataset diversity and emphasized the robustness of our models.

TABLE 3. Weighted average performance for consolidated violation categories

Category	Precision	Recall	F1	Support
Non-violating	0.73	0.86	0.79	42,267
Violating	0.59	0.66	0.63	28,174

TABLE 4. Performance of License Plate Detection (LPD) models on UFPR-ALPR dataset

LPD Model	Precision	Recall	F1 Score	mAP@0.5
Laroca *et al.* [5]	-	0.9833	-	-
YOLOv11n	0.9729	0.9789	0.9759	0.9910
YOLOv11m	0.9601	0.9719	0.9660	0.9846

D. LICENSE PLATE CHARACTER DETECTION (LPCD) MODELS

The performance of the LPCD model was evaluated on both synthetic and real datasets under different training regimes, including training solely on synthetic data, training solely on real data, and the proposed two-stage fine-tuning approach. Table 5 summarizes the quantitative results.

Training exclusively on synthetic data (Experiment 1) yielded excellent performance on synthetic validation images (mAP@0.5 = 0.9948), demonstrating the effectiveness of large-scale synthetic samples for learning character features. However, evaluation on real data (Experiment 2) revealed a notable performance drop, with precision decreasing by 7.90% and recall by 14.15%, highlighting the limitations of domain shift.

Using only real samples for training (Experiment 3) improved precision on real test data (+2.98% compared to Experiment 2) but slightly decreased recall (-3.35%), indicating that limited real data captures fewer variations. The two-stage approach (Experiment 4), where a synthetically trained model was fine-tuned on just 500 real samples, achieved the best results. Precision increased by 4.19% and recall by 9.22% over Experiment 2, while mAP@0.5 improved by 4.46% and mAP@[0.5–0.95] by 14.20% (0.8813 vs. 0.7717), confirming that synthetic pretraining provides transferable features, and modest real fine-tuning effectively bridges the domain gap.

The normalized vertical variation method reliably distinguished single-line and double-line plates, supporting accurate character sequencing. Overall, the LPCD pipeline achieved 98.46% reading accuracy on

TABLE 5. LPCD Model Performance Metrics under different training experiments

Experiment	Training Data	Validation / Test Data	Precision	Recall	mAP@0.5	mAP@[0.5–0.95]
1	Synthetic Only (PS-LPCD 32k)	Synthetic Validation (8k)	0.9948	0.9921	0.9948	0.9546
2	Synthetic Only (PS-LPCD 32k)	Real Test (PR-LPCD 150)	0.9162	0.8517	0.9375	0.7717
3	Real Only (PR-LPCD 500)	Real Test (PR-LPCD 150)	0.9435	0.8232	0.8803	0.7515
4	Synthetic + Fine-tune on Real (PSLPCD 32k + PR-LPCD 500)	Real Test (PR-LPCD 150)	0.9546	0.9302	0.9793	0.8813

the 650-image PRLPCD dataset. Failures occurred mainly under extreme rotations or shearing (Figure 7), where, for example, plate C missed the last character “7” as it fell below the midline, whereas plates A and B with milder distortions were read correctly as LEC-967-17.

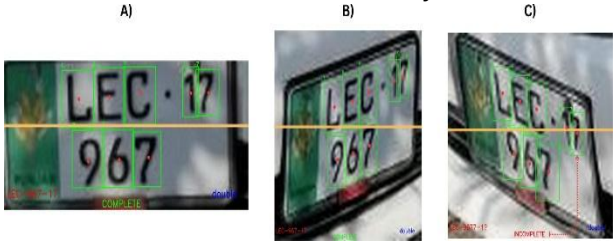

FIGURE 7. Examples of license plates demonstrating correct reading and failure cases.

Further analysis of the PR-LPCD real test set (150 plates) revealed that approximately 78% of recognition errors originated from the small-character regions inherent to Punjab license plate designs. These regions are particularly sensitive to adverse imaging conditions.

Typical failure cases include environmental degradation (LP_0019), where dust caused confusions such as ‘5’→‘S’ and ‘0’→‘U’; geometric distortion (LP_0332 and LP_0341) due to extreme viewing angles combined with low-resolution

small characters; physical wear (LP_0345), where faded printing reduced character contrast; and low-light conditions (LP_0401 and LP_0589), leading to misclassification among visually similar digits. These representative failures are illustrated in Figure 8.

The predominance of small-character-related errors suggests that robustness could be further improved through targeted synthetic augmentations (dust, blur, perspective warping) and enhanced multi-scale feature extraction.

Comparative evaluation on the 150-image PR-LPCD real test set further demonstrates the effectiveness of the proposed approach compared with EasyOCR [38] and PaddleOCR [39], the latter being based on a PP-OCR framework with a CRNN-style text recognition architecture, as shown in Table 6.

The proposed LPCD model consistently outperforms both general-purpose OCR baselines across all evaluation metrics. Compared to EasyOCR, character-level accuracy improves by 12.60% (from 87.67% to 98.72%), while fullplate match accuracy increases substantially from 33.33% to 90.67%. When compared with PaddleOCR, which demonstrates stronger character recognition performance (95.84%), the proposed method still achieves higher character-level accuracy and nearly doubles the full-plate match accuracy (49.33% to 90.67%).

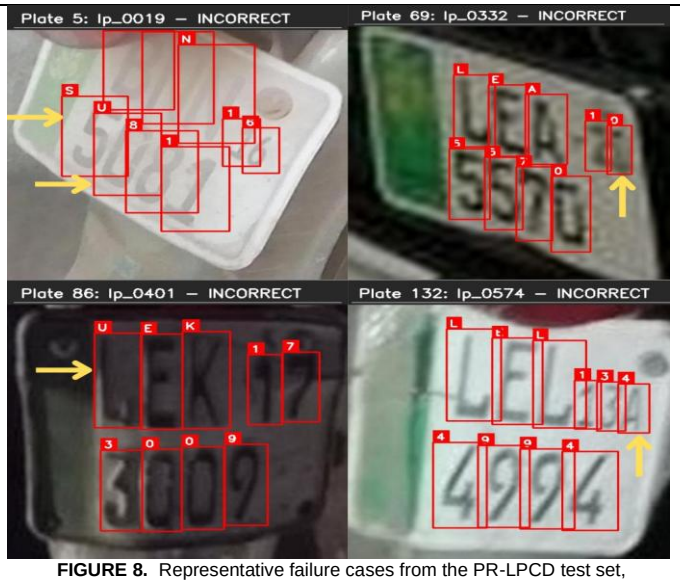

FIGURE 8. Representative failure cases from the PR-LPCD test set, highlighting typical character recognition errors. Yellow arrows indicate incorrectly predicted characters.

TABLE 6. LPCD vs. OCR Baselines Performance Comparison

Method	Character-level Accuracy	Full-Plate Match Accuracy
EasyOCR	0.8767	0.3333
PaddleOCR	0.9584	0.4933
LPCD Model	0.9872	0.9067

In addition to accuracy gains, the LPCD model demonstrates superior computational efficiency relative to EasyOCR. Total inference time on the test set is reduced by approximately 2.63× (from 53.04 s to 20.13 s), with average per-plate latency decreasing from 0.0816 s to 0.0310 s. These results demonstrate that task-specific pretraining on synthetic license plate data, combined with the proposed double-line classification heuristic, yields substantial improvements in recognition accuracy while maintaining real-time applicability in practical deployment scenarios.

E. END-TO-END PIPELINE INFERENCE

Two pipeline configurations integrating MCVD, LPD, and LPCD were evaluated on an NVIDIA RTX 2060 GPU:

- **NNN**: Nano variants for all models
- **MMN**: Medium variants for MCVD and LPD, Nano for LPCD

Each configuration was executed 5,000 times across 20 test images to estimate baseline inference speed without additional preprocessing. The NNN configuration achieved ~19 FPS, demonstrating real-time capability, while the MMN configuration achieved ~15 FPS, reflecting a trade-off between speed and accuracy.

To evaluate real-world performance, we applied the proposed sequential pipeline (see Section III-A) to five traffic surveillance videos from Lahore, Pakistan. These videos are available in our GitHub repository (https://github.com/mhdatheek136/P-LPCD). The results are summarized in Table 7.

TABLE 7. End-to-end pipeline inference results on traffic surveillance videos

Video	Pipeline FPS	Motorcycles Detected	Crowdedness
hd1	8.69	1596	1.93
hd2	8.11	3147	1.78
hd3	6.97	4181	2.78
hd4	9.17	1633	1.22
hd5	6.89	4725	2.67

The results show a strong dependency between crowdedness (average motorcycles per frame) and inference speed. With approximately one motorcycle per frame (e.g., Video hd4, crowdedness = 1.22), the pipeline maintained ~9 FPS. When crowdedness exceeded two motorcycles per frame (e.g., Videos hd3 and hd5), throughput dropped to ~7 FPS, revealing a near-linear decline in performance as object density increased.

Extrapolation suggests that under conditions of exactly one motorcycle per frame, the pipeline would sustain approximately 9.5 FPS. In contrast, at ~3 motorcycles per frame, throughput decreases by nearly 25–30%. This performance– density trade-off highlights the importance of optimizing inference strategies for deployment in dense urban environments.

V. CONCLUSION AND FUTURE WORK

The paper introduces a full YOLOv11-based motorcycle traffic violation detection pipeline, in particular, dealing with multi-rider detection, helmet-usage detection, and automatic license plate recognition (ALPR). The system is designed to suit Punjab, Pakistan, where motorcycles are a major cause of road deaths, and the levels of helmet usage among motorcycles are very low.

Key contributions of this study include the consolidation of classes for the MCVD model, achieving *mAP@50* scores of 0.66 with YOLOv11n and 0.71 with YOLOv11m, as well as strong license plate detection performance with *mAP@50* values approaching 0.99. The proposed PS-LPCD synthetic dataset, fine-tuned with the PR-LPCD real-world dataset, achieved 0.98 accuracy on real test sets. Both datasets (PSLPCD and PR-LPCD) are publicly available to support future research and to provide a reproducible procedure for creating region-specific license plate datasets, which is particularly useful in scenarios where real-world data is limited. Moreover, the normalized vertical variation technique that was proposed to differentiate between single-line and double-line plates is a lightweight alternative to the traditional deep learning methods. To provide a powerful video-based analysis, the BoT-SORT tracker was added, and the presented sequential pipeline could maintain around 9.5 FPS when it was applied to the scenes with one motorcycle per frame.

Several challenges were mitigated using task-specific augmentations, particularly the lack of near-camera close-up views in CCTV footage, which results in small and partially occluded license plates, along with varied camera perspectives and limited training data for multi-passenger cases. However, some limitations remain, such as lower efficiency when processing full-resolution images, potential identity mismatches in congested traffic, reliance on continuously updated synthetic datasets to accommodate evolving license plate formats, and difficulty in detecting small character regions on Punjab license plates under extreme conditions.

The following directions are suggested for future research:

(i) the use of regions of interest (ROIs) and rider-passenger relations modeling to enhance efficiency; (ii) the separation of nearby riders by the use of segmentation or pose estimation; (iii) the use of lightweight language models to refine the predictions of character sequences in the license plates;(iv) Controlled ablation studies to assess the impact of synthetic close-up augmentation on MCVD performance and the influence of synthetic data volume on LPCD accuracy; (v) parallelization of the ALPR process to enhance inference speed in a real-world application; and (vi) Developing specialized detection strategies for small character regions, particularly on Punjab license plates, to improve robustness under challenging conditions. These enhancements are meant to improve the strength, accuracy, and scalability of the proposed system to be used in large-scale traffic monitoring and enforcement applications.

ACKNOWLEDGMENT

We thank the contributors of the HELMET dataset [4], PK Number Plates V3 [32], and UFPR-ALPR dataset [5] for making their data publicly available for research purposes.

REFERENCES

[1] World Health Organization, *Global status report on road safety 2023: Summary*. Geneva, Switzerland: WHO, 2023. ISBN: 978-92-4-008645-6. [Online]. Available: https://www.who.int/publications/i/item/9789240086456.Accessed on: Sept. 18, 2025.

[2] A. Pervez, J. Lee, and H. Huang, "Identifying factors contributing to the motorcycle crash severity in Pakistan," *Journal of Advanced Transportation*, vol. 2021, Art. no. 6636130, 2021, DOI: 10.1155/2021/6636130.

[3] E. A. Mallhi, A. Sharif, and R. H. Javaid, "Pattern and frequency of motorcycle accident injuries reported to CMH Pannu Aqil, Sind," *Pakistan Armed Forces Medical Journal*, vol. 67, no. Suppl-1, pp. S58–S63, 2017.

[4] H. Lin, J. D. Deng, D. Albers, and F. W. Siebert, "Helmet Use Detection of Tracked Motorcycles Using CNN-Based Multi-Task Learning," *IEEE Access*, vol. 8, pp. 162073–162084, 2020, DOI: 10.1109/ACCESS.2020.3021357.

[5] R. Laroca, E. Severo, L. A. Zanlorensi, L. S. Oliveira, G. R. Gonçalves, W.R. Schwartz, and D. Menotti, "A robust real-time automatic license plate recognition based on the YOLO detector," in *Proc. 2018 Int. Joint Conf. Neural Netw. (IJCNN)*, 2018, pp. 1–10.

[6] J. Chiverton, "Helmet presence classification with motorcycle detection and tracking," *IET Intelligent Transport Syst.*, vol. 6, no. 3, pp. 259–269, 2012, DOI: 10.1049/iet-its.2011.0138.

[7] R. Waranusast, N. Bundon, V. Timtong, C. Tangnoi, and P. Pattanathaburt, "Machine vision techniques for motorcycle safety helmet detection," in *Proc. 2013 28th Int. Conf. Image and Vision Comput. New Zealand (IVCNZ)*, 2013, pp. 35–40, DOI: 10.1109/IVCNZ.2013.6726989.

[8] R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, and A. Soares, "Automatic detection of motorcyclists without helmet," in *Proc. 2013 XXXIX Latin American Comput. Conf. (CLEI)*,

2013, pp. 1–7, DOI: 10.1109/CLEI.2013.6670613.

[9] A. Hirota, N. H. Tiep, L. V. Khanh, and N. Oka, "Classifying Helmeted and Non-helmeted Motorcyclists," in F. Cong, A. Leung, and Q. Wei, Eds., *Advances in Neural Networks ISNN 2017*. Cham: Springer International Publishing, 2017, pp. 81–86, DOI: 10.1007/978-3-319-59072-1_10.

[10] M. Dasgupta, O. Bandyopadhyay, and S. Chatterji, "Automated Helmet Detection for Multiple Motorcycle Riders using CNN," in *Proc. 2019 IEEE Conf. Inf. Commun. Technol.*, 2019, pp. 1–4, DOI: 10.1109/CICT48419.2019.9066191.

[11] C. A. Rohith, S. A. Nair, P. S. Nair, S. Alphonsa, and N. P. John, "An Efficient Helmet Detection for MVD using Deep learning," in *Proc. 2019 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, 2019, pp. 282–286, DOI: 10.1109/ICOEI.2019.8862543.

[12] F. W. Siebert and H. Lin, "Detecting motorcycle helmet use with deep learning," *Accident Anal. Prevent.*, vol. 134, Art. no. 105319, 2020, DOI: 10.1016/j.aap.2019.105319. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S0001457519308401. Accessed on: Sept. 18, 2025.

[13] A. Chairat, M. Dailey, S. Limsoonthrakul, M. Ekpanyapong, and D. R. KC, "Low cost, high performance automatic motorcycle helmet violation detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, 2020, pp. 3560–3568.

[14] N. Kumar, G. K. Sahu, M. Ravi, S. Kumar, V. Sukruth, and A. N. M. Rao, "Triple Riding and No-Helmet Detection," in *Proc. 2023 4th IEEE Global Conf. for Advancement in Technology (GCAT)*, 2023, pp. 1–6, DOI: 10.1109/GCAT59970.2023.10353440.

[15] A. Goyal, D. Agarwal, A. Subramanian, C. V. Jawahar, R. K. Sarvadevabhatla, and R. Saluja, "Detecting, tracking and counting motorcycle rider traffic violations on unconstrained roads," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 4303–4312.

[16] T. Waris, M. Asif, M. B. Ahmad, T. Mahmood, S. Zafar, M. Shah, and A. Ayaz, "CNN-Based Automatic Helmet Violation Detection of Motorcyclists for an Intelligent Transportation System," *Mathematical Problems in Engineering*, vol. 2022, Art. no. 8246776, 2022, DOI: 10.1155/2022/8246776.

[17] A. Afzal, H. U. Draz, M. Z. Khan, and M. U. G. Khan, "Automatic helmet violation detection of motorcyclists from surveillance videos using deep learning approaches of computer vision," in *Proc. 2021 Int. Conf. Artif. Intell. (ICAI)*, 2021, pp. 252–257, DOI: 10.1109/ICAI52203.2021.9445206.

[18] M. Usama, H. Anwar, and S. Anwar, "Vehicle and license plate recognition with novel dataset for toll collection," *Pattern Anal. Appl.*, vol. 28, pp. 57, 2025, DOI: 10.1007/s10044-025-01443-8.

[19] S. U. Rehman, M. Ahmad, A. Nawaz, and T. Ali, "An efficient approach for vehicle number plate recognition in Pakistan," *The Open Artificial Intelligence Journal*, vol. 6, no. 1, pp. 12–21, 2020.

[20] J. Mistry, A. K. Misraa, M. Agarwal, A. Vyas, V. M. Chudasama, and K. P. Upla, "An automatic detection of helmeted and non-helmeted motorcyclist with license plate extraction using convolutional neural network," in *Proc. 2017 Seventh Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, 2017, pp. 1–6, DOI: 10.1109/IPTA.2017.8310092.

[21] L. Allamki, M. Panchakshari, A. Sateesha, and K. S. Pratheek, "Helmet detection using machine learning and automatic License Plate Recognition," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 6, no. 12, pp. 4475–4480, 2019.

[22] R. G. Nandhakumar, K. N. Devi, N. Krishnamoorthy, S. Shanthi, V. R. Pranesh, and S. Nikhalyaa, "Yolo Based License Plate Detection Of Triple Riders And Violators," in *Proc. 2023 Int. Conf. Comput. Commun. Informat. (ICCCI)*, 2023, pp. 1–6, DOI: 10.1109/ICCCI56745.2023.10128537.

[23] D. Anusha, D. S. Matta, D. A. Kumar, and D. Prabhakar, "Identifying Bikers Without Helmets and Triple Riding Automatically," in *Cognitive Computing and Cyber Physical Systems*. Cham: Springer Nature Switzerland, 2025, pp. 464–473, DOI: 10.1007/978-3-031-77081-4_36.

[24] R. R. Muley, G. K. Kethura, S. Lalitha, Y. Battula, and T. K. Sri, "Intelligent Traffic Monitoring System Using Deep Learning: Triple Riding, Automatic License Plate Recognition, and Helmet Detection," in *Proc. Fifth Doctoral Symp. Comput. Intell.*, Singapore: Springer Nature Singapore, 2024, pp. 391–402, DOI: 10.1007/978-981-97-6726-7_31.

[25] G. Silvano, V. Ribeiro, V. Greati, et al., "Synthetic image generation for training deep learning-based automated license plate recognition systems on the Brazilian Mercosur standard," *Designs, Automation & Embedded Systems*, vol. 25, pp. 113–133, 2021, DOI: 10.1007/s10617-020-09241-7.

[26] A. Spruck, M. Gruber, A. Maier, D. Moussa, J. Seiler, C. Riess, and A. Kaup, "Synthesizing Annotated Image and Video Data Using a Rendering-Based Pipeline for Improved License Plate Recognition," *arXiv preprint arXiv:2209.14448*, 2022. [Online]. Available: https://arxiv.org/abs/2209.14448. Accessed on: Sept. 18, 2025.

[27] M. Shpir, N. Shvai, and A. Nakib, "License Plate Images Generation with Diffusion Models," *arXiv preprint arXiv:2501.03374*, 2025. [Online]. Available: https://arxiv.org/abs/2501.03374. Accessed on: Sept. 18, 2025.

[28] M. Del Castillo Velarde and G. Velarde, "Benchmarking Algorithms for Automatic License Plate Recognition," *arXiv preprint arXiv:2203.14298*, 2022. [Online]. Available: https://arxiv.org/abs/2203.14298. Accessed on: Sept. 18, 2025.

[29] Ultralytics, "Yolo11 new," 2025. [Online]. Available: https://docs.ultralytics.com/models/yolo11/. Accessed on: Jan. 26, 2025.

[30] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "BoT-SORT: Robust associations multi-pedestrian tracking," *arXiv preprint arXiv:2206.14651*, 2022. [Online]. Available: https://arxiv.org/abs/2206.14651. Accessed on: Sept. 18, 2025.

[31] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training Real-World Blind Super-Resolution With Pure Synthetic Data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops*, Oct. 2021, pp. 1905–1914.

[32] B. Khan, "PK-Number-Plates-V3," Roboflow Universe, 2023. [Online]. Available: https://universe.roboflow.com/burhan-khan/pk-number-plates. License: CC BY 4.0. Accessed on: Sept. 18, 2025.

[33] A. Mohamed Rafi, F. A. F. Mohamed Buhary, and A. Ishaq Khan, "Punjab Pakistan Synthetic and Real License Plate Character Datasets (P-LPCD)," *Zenodo*, ver. 1.0.0, Sept. 23, 2025, DOI: 10.5281/zenodo.17182320.

[34] R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv:2410.17725*, 2024. [Online]. Available: https://arxiv.org/abs/2410.17725. Accessed on: Sept. 18, 2025.

[35] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014, pp. 740–755, DOI: 10.1007/978-3-319-10602-1_48.

[36] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.

[37] N. S. Keskar and R. Socher, "Improving generalization performance by switching from adam to sgd," *arXiv preprint arXiv:1712.07628*, 2017. [Online]. Available: https://arxiv.org/abs/1712.07628. Accessed on: Sept. 18, 2025.

[38] JaidedAI, *EasyOCR: Ready-to-Use Optical Character Recognition*, GitHub repository, 2021. [Online]. Available: https://github.com/JaidedAI/EasyOCR.

[39] Y. Du, C. Li, R. Guo, S. Xu, Y. Liu, J. Wang, B. Zhou, and X. Bai, "PaddleOCR: A Lightweight and Efficient OCR System," *arXiv preprint arXiv:2009.09941*, 2020. [Online]. Available: https://arxiv.org/abs/2009.09941. Accessed on: Sept. 18, 2025.

APPENDIX A CLASS DISTRIBUTION BEFORE AND AFTER CONSOLIDATION OF THE HELMET DATASET

This appendix shows the HELMET dataset's class distribution before and after consolidation. Initially comprising 36 fine-grained classes (Figure 9), the dataset was merged and augmented into 7 broader categories (Figure 10) to reduce class imbalance.

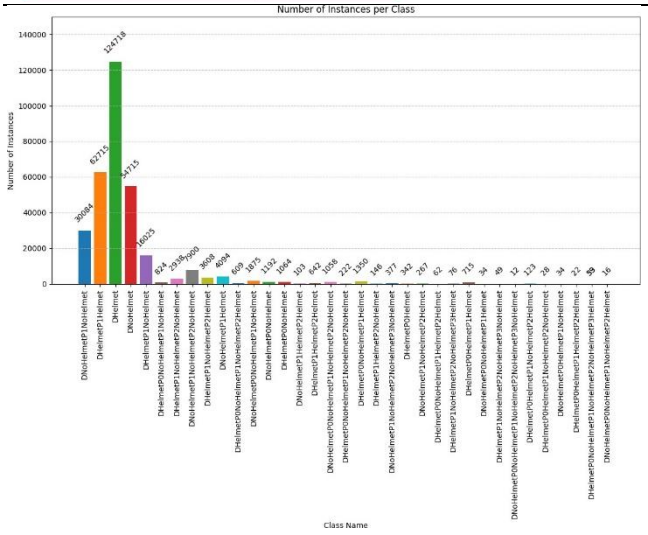

FIGURE 9. Class distribution of the HELMET dataset before consolidation (36 classes).

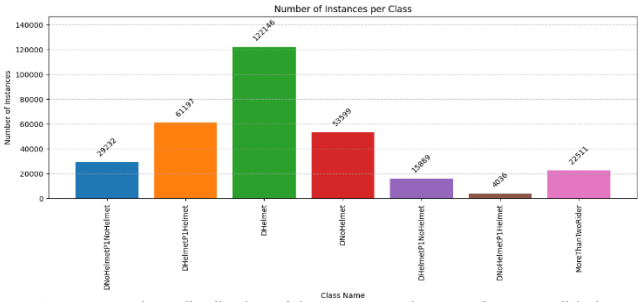

FIGURE 10. Class distribution of the HELMET dataset after consolidation and augmentation (7 classes).

APPENDIX B TRAINING AND VALIDATION CURVES

This appendix provides the full training dynamics for all YOLOv11 models used in MCVD, LPD, and LPCD experiments. Each figure shows training and validation loss curves (box, cls, dfl) along with validation mAP, precision, and recall over epochs.

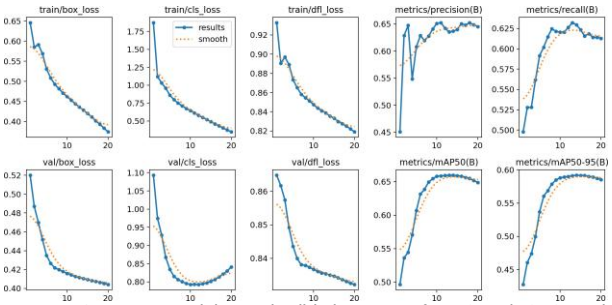

FIGURE 11. Training and validation curves for MCVD (YOLOv11n).

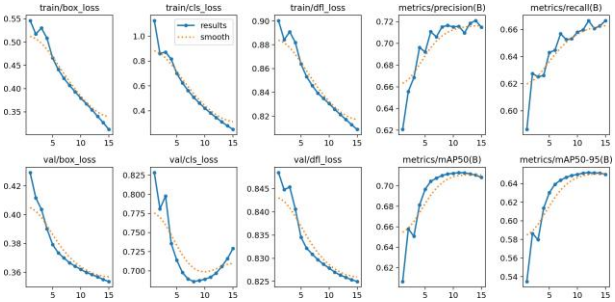

FIGURE 12. Training and validation curves for MCVD (YOLOv11m).

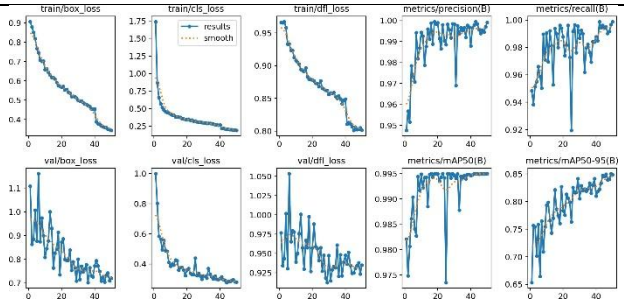

FIGURE 13. Training and validation curves for LPD (YOLOv11n).

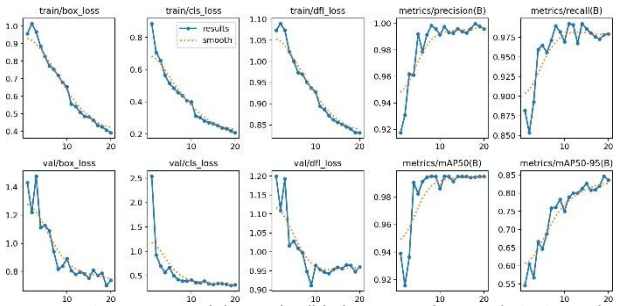

FIGURE 14. Training and validation curves for LPD (YOLOv11m).

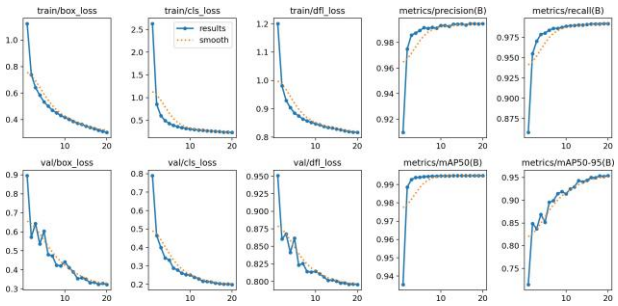

FIGURE 15. Training and validation curves for LPCD (YOLOv11n).

Date available online: 09-01-2026
Vol. 3, Issue 2 (July – December 2025)

DOI: https://doi.org/10.24312/ucp-jeit.03.02.765

Optimizing Breast Cancer Screening Outcomes in Dense Breasts: Diagnostic Performance, Cost-Effectiveness, and Implementation of Supplemental Modalities

Malaika Arif[1], Imran Ahmad[2], Fatima Abbas[1], Sunbal Faraz Hayat[3]
[1]International Collaborative Research Group, Lahore, Pakistan
[2]Riphah International University, Malakand Campus, Lower Dir, Pakistan
[3]Pakistan Navy, Islamabad, Pakistan

Corresponding Author: Imran Ahmad (email: imran.ahmad@riphah.edu.pk)

Abstract

Dense breast tissue is a major challenge to screening of breast cancer since it covers about 40-50% of the breast and hides the lesions in traditional mammography. This weakness requires the efficient use of additional imaging modalities for high-risk patients. The following paper is a comparative analysis of the major supplemental imaging techniques, such as ultrasound, Magnetic Resonance Imaging (MRI), Contrast-Enhanced Spectral Mammography (CESM), Digital Breast Tomosynthesis (DBT), and AI-assisted mammography. A literature review was conducted in databases like PubMed, ScienceDirect, and Google Scholar, and peer-reviewed articles published since 2016, and those that concentrated on breast cancer screening in high-density breasts were assessed. The review is a confirmation of the role that mammography plays at its core, but indicates the improved detection of cancer with the help of supplemental ultrasound and MRI. Importantly, CESM has a similar diagnostic potential as MRI and the practical advantages of lower costs and shorter scan time. DBT enhances clear images by reducing overlap between tissues, and it effectively reduces the rate of recollection among patients. Moreover, AI-assisted mammography is one of the essential developments that will raise the detection of cancer in dense breasts and may even prevent the use of auxiliary procedures. The results strongly reflect the need of implementation of tailored and risk-based screening methods where such advanced add-on imaging technologies are combined with Artificial Intelligence, which is likely to elevate screening accuracy, clinical outcomes, affordability, and ultimately make a significant contribution to lowering the rates of mortality of breast cancer.

INDEX TERMS: Breast Density, Breast Cancer Screening, Supplemental Screening Strategies, Breast Cancer Risk, Ultrasound, MRI, CESM, DBT, AI, Risk-Stratified Screening, Diagnostic Performance, Cost-Effectiveness, MRI (AB-MRI)

I. INTRODUCTION

Well-planned screening helps to increase the survival rates of breast cancer. It remains the most widespread type of cancer in women worldwide, and early diagnosis has a considerable impact on the results. One of the determinants that influences the risk of cancer and the accuracy of the screening process is the breast density, or the percentage of fat to fibroglandular tissue that is seen on the mammogram. It is also advised that if women go through mammographic screening annually, the chances of death by breast cancer are minimized by 20 percent [1]. As suggested by previous studies, around 40-50% of women over the age of 40 have dense or heterogeneous breasts, which are further classified as C and D in the Breast Imaging Reporting and Data System (BI-RADS) [2].

The masking effect of highly dense breast tissue can significantly reduce the sensitivity of conventional mammography, hindering early cancer detection [3]. Supplemental modalities, such as contrast-enhanced spectral mammography (CESM), have been investigated. Meta-analyses indicate that CESM provides high specificity and sensitivity in identifying lesions hidden in dense breasts [4]. Magnetic resonance imaging is also one of the most sensitive imaging techniques; a recent study evaluated MRI and found its sensitivity and specificity to be higher than that of mammography [5]. The accuracy, specificity, and recall rate of mammography with supplemental ultrasonography were higher than those of mammography alone [6]. Digital breast tomosynthesis (DBT) offers superior lesion visualization than conventional mammography, as suggested by certain previous studies. However, its advantage for women with highly dense breasts is still unclear [7]. To improve early detection and provide the best individualized treatment, it is necessary to determine the best screening method for women with dense breasts [1].

According to previous studies, millions of women are diagnosed with breast cancer each year. It is also noted that breast cancer continues to be one of the leading causes of female mortality, with hundreds of thousands of them dying from the disease annually [8]. Various

screening guidelines are employed in order to diagnose breast cancer at an early stage. The protocols enhance survival rates during breast cancer patients with dense breast since they provide less invasive treatments [9]. Conventional mammography is still the most common screening approach available because it is easily accessible and can effectively reduce the rate of mortality [10]. It is very insensitive in women who have dense breasts and brings about a high risk of cancer by hiding lesions and abnormalities [11]. Researchers are investigating personalized screening by taking individual risk factors and breast density into account for detecting breast cancer early [12]. Other screening tools, including ultrasound, CESM, MRI, and AI mammography, are being explored because mammography doesn't function well for dense breasts [6, 13]. Physicians get help from these techniques in customizing screening to each woman's unique requirements [1].

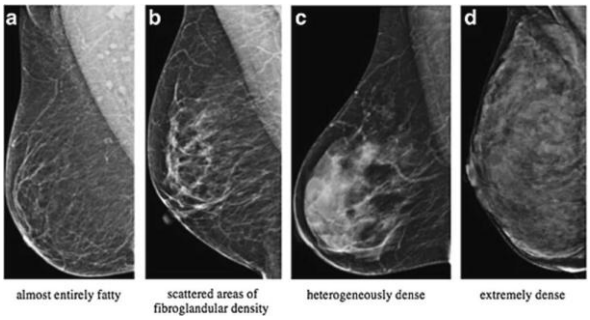

Figure 1: BI-RADS breast density categories shown from low to high density [3].

The proportion of fibroglandular tissue to fatty tissue seen on a mammogram is known as breast density [11]. Breast density is divided into four categories based on the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS): A–almost entirely fatty, B–diffuse fibroglandular, C–heterogeneously dense, and D–extremely dense [14]. Dense breasts (categories C or D) not only make tumors harder to detect on mammography but also increase the breast cancer risk [3]. The phenomenon "masking effect" makes abnormalities harder to detect because both tumors and dense tissue appear white on mammograms. It is common practice to use additional imaging techniques to address the limitations of mammography. Reconstructions of three-dimensional images used in digital breast tomosynthesis can improve lesion visibility and reduce recall rates [7]. In terms of diagnostic accuracy comparable to that of MRI, Contrast-Enhanced Spectral Mammography (CESM) allows vascular assessment of lesions by combining low- and high-energy imaging [4]. Ultrasound (US) is a safe imaging method because it does not use radiation and can detect tumors in women with dense breasts that mammography might miss, despite being dependent on the operator's skill [15]. Magnetic resonance imaging (MRI), because of its high cost and restricted availability, is still the most sensitive method for detecting breast cancer, especially in women who are at high risk [5]. Sensitivity indicates how well the test detects women who actually have the disease, while specificity refers to how well the test detects women who do not have the disease. High specificity minimizes false positives and unnecessary follow-ups, while high sensitivity allows for early identification, particularly in women with dense breasts. Such definitions align with previous research assessing breast density, cancer risk, and screening performance [4].

Supplemental imaging techniques (MRI, CESM, DBT, and ultrasound) have been independently evaluated for breast screening in women with high breast density. During the past few years, several large-scale trials and systematic reviews have shown improvements in detection rates over mammography alone [13]. Studies indicate that in many dense-breast cases, contrast-enhanced spectral mammography (CESM) can achieve sensitivity similar to MRI; its specificity and cost vary significantly among healthcare settings [16]. In breasts

Table 1. PICOS framework for inclusion and exclusion criteria

Component	Inclusion Criteria	Exclusion Criteria
Population (P)	Women with heterogeneously or extremely dense breast tissue and negative mammography results	Studies involving non-dense breasts, male patients, or non-human studies
Intervention/Exposure (I)	Use of supplemental imaging techniques, such as contrast-enhanced spectral mammography (CESM), digital breast tomosynthesis (DBT), magnetic resonance imaging (MRI), and ultrasound (US)	Studies limited to standard 2D mammography
Comparison (C)	Compared with standard 2D mammography or between different supplemental techniques	Studies without a reference standard
Outcomes (O)	Diagnostic performance matrices that are reported include cost-effectiveness, sensitivity, specificity, recall rate, PPV, and cancer detection rate (CDR)	Studies lacking diagnostic outcomes
Study Design (S)	Prospective or large retrospective cohort studies, cross-sectional studies, and meta-analyses of diagnostic accuracy (2016-2025)	Single-case reports, editorials, and small-sample studies
Other Criteria	Peer-reviewed English publications between 2016 and 2025	Non-peer-reviewed publications in other languages or before 2016

with different densities, digital breast tomosynthesis (DBT) has been shown to reduce recall rates and enhance lesion visibility; however, the further advantage for highly dense tissue is still less certain [7]. AI-assisted mammography can also be used to eliminate the need for additional mammography, with sensitivity and specificity comparable to those of supplemental ultrasound. Furthermore, these advancements from standard screening methods to personalized techniques mainly focus on the early detection and prevention of breast cancer [6]. Recent studies also show that using several screening modalities without specific guidelines for when and how often to apply those increases the number of false positives and incorrect diagnoses [6]. Few studies have comprehensively evaluated key supplemental imaging modalities across different settings and dense-breast populations to establish an optimal balance between their benefits and risks [17]. The absence of comprehensive long-term outcome data, including interval cancer incidence, stage at diagnosis, and mortality, limits understanding of its real-world impact [2]. Furthermore, many proposals may not be practical worldwide because the limited resources of low- and middle-income regions were not properly taken into consideration [12]. To incorporate the most recent studies, compare diagnostic performance, analyze risks, and assess feasibility across settings for women with dense breasts, a thorough synthesis is necessary. This review assesses the diagnostic efficiency and cost-effectiveness of supplemental screening modalities, including digital breast tomosynthesis (DBT), ultrasound (US), magnetic resonance imaging (MRI), contrast-enhanced spectral mammography (CESM), and AI in women with dense breasts and negative mammogram reports from previous studies. It also looks for research gaps to provide risk-based ideal screening methods.

Research Questions (RQs):

RQ1: What are the current challenges in breast density assessment?

RQ2: How do supplemental imaging modalities compare in detecting breast cancer in dense breasts?

RQ3: In what ways do several imaging modalities affect diagnostic performance?

II. METHODOLOGY

The central focus of this review was on systematically examining supplemental screening techniques for women with an elevated risk of breast cancer who have dense breast tissue. In accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) principles, the review was conducted, however, no quantitative synthesis was performed, and the review was not registered prospectively in PROSPERO. Further, a transparent and organized methodology was applied to ensure integrity throughout the review process.

A. Search Strategy

Information Sources: A comprehensive search was conducted to identify recent papers and relevant studies published between August 2016 and July 2025. PubMed, PubMed Central (PMC), ScienceDirect, and Google Scholar from widely accessible digital libraries were used. Specialized databases such as Embase and Cochrane were not part of the search strategy due to a lack of access and resource limitations. All search decisions and limitations were explicitly documented to maintain transparency.

Search Query: To identify up-to-date research, a comprehensive search was carried out by exploring selected databases. Medical Subject Headings (MeSH) and keyword terms such as “dense breasts”, “supplemental imaging techniques”, “contrast-enhanced spectral mammography”, “digital breast tomosynthesis”, ultrasound”, “magnetic resonance imaging”, and artificial intelligence” were used in the search strategy.

Boolean operators AND, OR, and NOT were used to narrow down the search results. This combination aided in the screening of human subject-focused research for high-risk women with dense breasts. The search method utilized the following aspects of a structured Boolean expression: (“dense breast” OR “breast density”) AND (“breast cancer screening” OR “screening mammography”) AND (“ultrasound” OR “magnetic resonance imaging” OR “MRI” OR “digital breast tomosynthesis” OR “contrast-enhanced spectral mammography” OR “CESM”) NOT (“animal” OR “case report”).

Other Search Methods: To ensure comprehensiveness, backward and forward snowballing were employed to review the reference lists of the included publications and related reviews. This review does not include grey literature, such as abstract-only research, conference proceedings, and other non-peer-reviewed data.

B. Eligibility (Inclusion/Exclusion) Criteria:

In accordance with the population, intervention, comparison, outcomes, and study design (PICOS framework), which is outlined in this table, articles were screened using specified inclusion and exclusion criteria:

C. Study Selection Process:

The selection of the study complied with the transparency and reproducibility requirements of the PRISMA guidelines. EndNote software was used to import references from PubMed, Google Scholar, ScienceDirect, and PMC for reference management and to prevent duplication. In the studies, screening was conducted in two stages. In the first step, all titles and abstracts were reviewed by two independent reviewers using PICOS-based eligibility criteria. When any differences could not be resolved through discussion, a third reviewer was brought in to make the final decision. A second round of review was conducted by the same reviewers who independently examined the full texts of all possibly qualifying articles. At the full-text stage, reasons for exclusion were carefully noted. To maintain uniformity and transparency, both stages were screened using the same criteria. In the final synthesis, only studies that met all inclusion criteria after reviewer agreement were included.

PRISMA flow diagrams were created to describe the process of identifying, screening, determining eligibility, and including participants. Using database searches and supplementary sources, 100 records were discovered, of which 90 were found through database searches and 10 by searching supplementary sources. There were only 30 records left for title and abstract screening after the duplicates were removed. As a result of this stage, 70 records were excluded because they did not meet the eligibility requirements. In total, 22 full-text publications were reviewed. Among these publications, 8 were removed due to insufficient methodological relevance or the absence of the complete text. There were 22 papers that met all criteria for inclusion in the final qualitative synthesis. As shown in Figure 2, the PRISMA flow diagram provides a visual representation of the numerical breakdown of each stage in order to ensure methodological transparency and reproducibility.

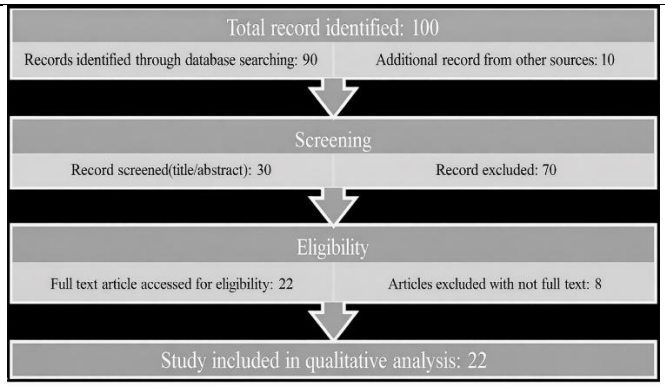

Figure 2: PRISMA flow diagram for study selection

The 22 included papers consist of 10 reviews or expert guidelines published between 2016 and 2025, 9 cohort studies published between 2016 and 2024, and 6 systematic reviews/meta-analyses published between

Table 2. Summary of included studies and their characteristics

Study Design	Number of Studies	Publication Years	References
Systematic Reviews / Meta-analyses	6	2021–2024	[2], [4], [7], [13], [16], [17]
Cohort Studies	7	2016–2024	[6], [11], [12], [14], [15], [19]-[22]
Reviews / Expert Guidelines	10	2016–2025	[1], [3], [5], [8], [9], [10], [18]
Total / Overall	22	2016–2025	[1]–[22]

Table 3. Summary of risk-of-bias assessment of the included studies.

Ref No.	Author Name	Study Design	Tool Used	Overall Risk of Bias
[1]	Mann et al.	Recommendation/Expert Guidelines	QUADAS-2	Moderate
[2]	Mokhtary et al.	Systematic Review & Meta-analysis	CASP	High
[3]	Nissan et al.	Narrative Review	CASP	Moderate
[4]	Liu et al.	Systematic Review & Meta-analysis	CASP	Low
[5]	Sitges et al.	Narrative Review	CASP	Moderate
[6]	Lee et al.	Retrospective Study	NIH	High
[7]	Raichand et al.	Systematic Review	CASP	Moderate
[8]	Bray et al.	Descriptive Epidemiological Study	QUADAS-2	Low
[9]	Marmot et al.	Independent Review	QUADAS-2	Moderate
[10]	Tomlinson-Hansen et al.	Narrative Review	NIH	High
[11]	Boyd et al.	Observational Cohort	NIH	Moderate
[12]	Bertsimas et al.	Cohort Observational Study	NIH	High
[13]	Abu Abeelh et al.	Systematic Review	CASP	High
[14]	Kim et al.	Cohort Study	NIH	Moderate
[15]	Vourtsis et al.	Cohort Observational Study	NIH	Low
[16]	Daniaux et al.	Systematic Review	CASP	Moderate
[17]	Tran et al.	Meta-analysis	CASP	Low
[18]	USPSTF et al.	Clinical practice guideline (USPSTF recommendation).	QUADAS-2	Moderate
[19]	Tan et al.	Retrospective observational study	NIH	Moderate
[20]	Mansour et al.	Retrospective observational study	CASP	High
[21]	Shermis et al.	Cohort Study	NIH	High
[22]	Richman et al.	Observational study	NIH	Moderate

2021 and 2024. Overall, the studies included a wide spectrum of evidence, from primary research to expert opinions. Table 2 describes a summary of the study's design, number of studies, years of publication, and references:

D. Data Extraction and Data Items

A structured Excel sheet was used to collect data from selected studies. It was mainly employed to accurately analyze the studies, ensuring accuracy, consistency, and comparability of the diagnostic performance of different supplemental modalities. The primary findings and limitations of these modalities were also recorded. A narrative synthesis was used to compare and summarize the diagnostic accuracy of different imaging techniques in an organized way. The data extraction strategy and synthesis protocols were used to accurately compare positive predictive value (PPV), recall rate, sensitivity, specificity, and cancer detection rate (CDR) from the available studies.

E. Risk of Bias (Quality) Assessment

This review carefully evaluated all included studies using different tools to assess their quality. The tools used included the QUADAS-2 tool, the CASP checklist, and the NIH Quality Assessment Tool were used to examine diagnostic accuracy studies, systematic reviews and meta-analyses, and observational studies, respectively. These tools were also used to evaluate methodological quality and possible risk of bias. After evaluation, all studies were categorized as having a low, moderate, or high risk of bias to guide the overall discussion. These ratings were then used to give greater importance to studies with stronger and more trustworthy methodologies. Table 3 summarizes the methodological characteristics of the included studies.

III. RESULTS

A. Synthesis of Results

The synthesis of results focuses on key aspects of breast screening in women with dense breast tissue. It further summarizes the results from 22 studies included in the review. Dense breasts not only increase the risk of breast cancer but also make it more difficult for mammography to detect abnormalities. MRI and CESM demonstrate the highest sensitivity among the imaging modalities assessed, whereas ultrasound and AI-assisted imaging provide supplementary support for identifying lesions that mammography might miss. Patient knowledge, medical recommendations, and accessibility substantially influence adherence. By combining multimodal screening methods, early detection may be enhanced, and the long-term outcome may be improved, permitting less aggressive treatment and reducing overall health-care costs. The diagnostic performance of several modalities was compared using a narrative synthesis. The following metrics are reported: sensitivity, specificity, positive predictive value (PPV), and cancer detection rate (CDR). Qualitative findings were supplied in cases where quantitative data were not available. As shown in Table 3, the narrative synthesis of diagnostic performance in different modalities was summarized.

Previous studies consistently identify several key challenges for the accurate assessment of breast density. Due to its ability to obscure lesions, increase false-negative results, and complicate early cancer detection, dense fibroglandular tissue lowers mammographic sensitivity [1], [11]. Variations in imaging methods and subjective interpretation of BI-RADS classification also affect breast density evaluation, resulting in uneven categorization among readers and institutions [6], [20]. Furthermore, single-point measures are not accurate for long-term risk assessment because breast density varies with age and hormonal factors [2], [14]. Volumetric and AI-based approaches still need to be validated before being used in clinical settings, considering their potential to standardize evaluation [19], [20].

Efficacy of Supplemental Screening Modalities:

A previous study showed that women who have very dense breasts need to have additional screening since they are likely to develop breast cancer and have lower mammography sensitivity. In women who are premenopausal or whose breast density changes quickly, mammography alone may miss malignancies in dense tissue [1], [2]. Combining ultrasound with mammography as an additional imaging modality increases detection rates, particularly for small and node-negative cancers [6]. Several studies suggests that breast MRI monitoring at longer intervals may be beneficial for women with thicker breasts, although the ideal frequency is still being researched [1]. Contrast-enhanced spectral mammography (CESM) offers high sensitivity and specificity and detects problems that traditional mammography may miss [4]. Digital breast tomosynthesis (DBT) increases the early detection rates of cancers in dense breast tissue. In contrast to conventional mammography, it also decreases recall rates [7]. AI-assisted imaging can be used in conjunction with 3D automated breast ultrasonography. It also reduces false negatives and increases early detection by enhancing mammographic analysis [19]. A woman with dense breasts who undergoes mammography sometimes shows a negative mammogram. To improve visibility and detection accuracy, molecular breast imaging provides a useful answer to the clinical problems of detecting cancer [21]. Personalized screening methods, which are designed around each patient's particular risk profile and tissue characteristics, are supported by previous studies. These methods aim to detect minor alterations sooner and provide more focused recommendations [12].

Predictors of Adherence to Screening Protocols:

Healthcare professionals provide essential recommendations that include practical considerations such as accessibility, cost, and insurance coverage. These factors also play a key role in determining a patient's involvement in supplemental imaging [13]. When women clearly understand the risks of dense breasts and the limitations of mammography in hiding

Table 3: A narrative synthesis of the diagnostic efficacy of screening methods for breast cancer in women with dense breasts

Screening Modalities	Sensitivity	Specificity	Cancer Detection Rate (CDR)	Positive Predictive Value (PPV)	References
2D Mammography	Sensitivity is lower due to dense tissue masking	Moderate specificity	CDR is lower in dense breasts	PPV is lower due to missed cancers	[8], [9], [11]
Digital Breast Tomosynthesis (DBT)	Improved sensitivity compared to 2D mammography	Slightly higher Specificity than 2D mammography	A moderate increase in CDR is observed	Improved PPV with reduced recall rates	[7], [22]
Ultrasound (HHUS/ABUS)	High sensitivity	Variable specificity	Improved CDR rate when combined with mammography	Moderate PPV; may decrease with increased false positives	[6], [15], [19]
Magnetic Resonance Imaging (MRI)	Higher sensitivity than other screening modalities (>90%)	Moderate specificity, approximately (70–85%)	Highest CDR among all modalities	High PPV, especially for invasive cancers	[1], [5], [13]
Contrast-Enhanced Spectral Mammography (CESM)	High sensitivity, approximately (85–90%)	Moderate specificity, approximately (75–85%)	CDR is Comparable to that of MRI; higher than 2D mammography	Moderate to high PPV	[3], [4], [16]
Molecular Breast Imaging (MBI)	High sensitivity, approximately (80%)	Moderate specificity, approximately (80–85%)	Detects additional cancers missed by mammography	Comparable to ultrasound; moderate PPV	[21]
AI-Based Multimodal Systems	High sensitivity, approximately (85%)	Maintains or slightly improves specificity	CDR improvement when integrated with mammography or ultrasound	Improved PPV by reducing missed lesions	[6], [19], [20]

most of the lesions, they are more likely to participate in supplemental screening programs other than mammography [3]. Public awareness campaigns and follow-up programs are organized. They have been effective in encouraging people's long-term adherence to recommended screening protocols [5].

Long-Term Outcomes and Cost-Effectiveness:

The cost-effectiveness in this section is discussed in a theoretical manner, because no quantitative economic study (ICER/QALY) has been performed. Supplemental screening techniques, especially MRI, CESM, and ultrasound, allow for early detection and can help reduce death rates when compared to mammography alone [1], [4]. Due to the higher initial costs of MRI and CESM, a problem arises that further necessitates the personalized, risk-based screening methods combining mammography, ultrasound, and AI-assisted imaging, which are cost-effective and reduce the incidence of interval cancers and long-term treatment expenses [12], [17], [19]. Using several methods further improves patient outcomes through early detection. It may reduce the need for aggressive therapies and allow for less invasive procedures, such as breast-conserving surgery rather than mastectomy [5], [16]. To maximize healthcare by improving diagnostic accuracy and reducing needless treatment and follow-ups is ensured by multimodal imaging tailored to each patient's risk profile, which maximizes diagnostic accuracy and minimizes needless treatment and follow-ups [7], [21]. Continuous adherence to supplemental screening is necessary for achieving such therapeutic advantages since non-compliance can compromise the ability of these strategies in lowering mortality and medical expenses [18].

B. Risk of Bias within Studies

The majority of the 22 studies were rated as being of moderate to high quality and made up the basis of this review. The high-quality studies with strong methodological rigor were systematic reviews and meta-analyses as demonstrated by thorough literature searches and consistent reporting of their findings [2], [4], [7], [13], [17]. Most cohort studies were prospective and had explicit inclusion criteria. However, several had small sample numbers or insufficient follow-up data [6], [14]-[16], [19], [21], [22]. Although reviews and expert suggestions were comprehensive, they were sometimes constrained by unclear search methodologies and varied

Table 4: Comparison of key studies that assessed various methods for screening for breast cancer in women with dense breasts.

Study (Ref.)	Population/ Setting	Modality/ Intervention	Main Findings	Strengths	Limitations
Liu J et al. [4]	Women getting their breast evaluation	Contrast-Enhanced Spectral Mammography (CESM)	CESM demonstrates good sensitivity and specificity comparable to MRI.	Large data pool; efficient meta-analysis.	Varying protocols; possible bias.
Sitges C & Mann RM [5]	Extremely dense breasts in women	Breast Magnetic Resonance Imaging (MRI) screening	MRI increased cancer detection with a moderate recall rate.	Concentrated on the dense-breast group; updated data.	High cost; limited availability.
Lee SE et al. [6]	Women with dense breasts	Mammography vs. Artificial Intelligence vs. Ultrasound	AI and US enhanced detection over mammography; AI accuracy was comparable to that of radiologists.	Direct tools comparison: practical significance.	Retrospective; small sample.
Raichand S et al. [7]	Dense-breast women with added risk factors	Digital Breast Tomosynthesis (DBT)	Comparing DBT to 2D mammography, the first method increased detection but reduced recalls.	Extensive review; multiple populations included.	No long-term outcomes.
Abu Abeelh E & AbuAbeileh Z [13]	Women with dense breast tissue	Mammography, Ultrasound, MRI	MRI was the most sensitive, followed by US and mammography	Clear comparative synthesis.	Few studies; varied methods.
Daniaux M et al. [16]	Newly diagnosed breast cancer patients	Contrast-enhanced Spectral Mammography(CESM) vs Mammography, US, MRI	CESM accuracy is similar to MRI and more accurate than US or mammography	Detailed multimodal comparison.	Focused on staging, not screening.
Tran E & Ray K [17]	Dense-breast women with negative mammograms	Meta-analysis of MRI, US, MBI	MRI works best; US and MBI have limited value.	Dense-breast subgroup; pooled analysis.	Study heterogeneity; no mortality data.
Shermis RB et al. [21]	Dense-breast women with negative mammography	Molecular Breast Imaging (MBI)	MBI detected ~7.7 extra cancers/1,000; recall 8.4%.	Real-world clinical data.	Retrospective: radiation exposure.

study populations [1], [3], [5], [8], [12], [18], [20]. Common methodological limitations in all of the included studies were variability in breast density classification, study population heterogeneity, variations in imaging procedures, and limited sample sizes in certain cohort studies.

IV. DISCUSSION

A. Summary of Evidence

This review evaluated breast cancer screening techniques for women with dense breast tissue. This study combined data from 22 studies. Mammographic breast density was consistently found to be a significant limitation of mammography due to its lower sensitivity and a strong independent risk factor for breast cancer across all studies [1], [11], [14]. In comparison to mammography alone, other imaging modalities, including digital breast tomosynthesis (DBT), contrast-enhanced spectral mammography (CESM), magnetic resonance imaging (MRI), ultrasound (US), molecular breast imaging (MBI), and artificial intelligence AI-assisted tools, showed better detection rates in dense breasts [4]-[7], [13], [16], [17], [21]. The highest sensitivity was obtained by MRI and CESM, with MRI exhibiting better lesion characterization and CESM emerging as a viable substitute in situations

Table 5: Comparative summary of cost and accessibility of breast cancer screening modalities in women with dense breasts.

Modality	Cost / Accessibility	Key References
2D Mammography	Lowest cost and most widely available modality; it forms the foundation of national screening programs worldwide.	[1], [18]
Digital Breast Tomosynthesis (DBT)	Moderately higher cost than 2D but increasingly available; compatible with existing mammography systems and feasible for large-scale screening.	[7]
Ultrasound (HHUS / ABUS)	Low to moderate cost; handheld ultrasound is widely available but operator-dependent, while automated systems improve standardization but require dedicated equipment.	[13]
Contrast-Enhanced Spectral Mammography (CESM)	Moderate cost; more affordable and accessible than MRI, requiring IV contrast but using standard mammography infrastructure.	[4]
MRI	Highest cost and limited accessibility; requires advanced equipment, longer examination time, and specialized interpretation—best suited for high-risk women.	[5]
AI-Assisted Imaging	Implementation cost remains variable, but integration improves efficiency and workflow. Accessibility is expanding with digital infrastructure and validation studies.	[6], [19]

where MRI availability or cost is a barrier [4], [5], [16]. Particularly in moderate-density categories, DBT and ultrasound offered progressive cancer detection [6], [7], [15]. Diagnostic accuracy was improved and false positives were decreased with the use of AI and multimodal techniques that integrated mammography with US or digital Breast Tomosynthesis [19], [20]. Overall, the results were consistent with previous studies and new worldwide screening guidelines that emphasize multimodal and risk-stratified screening for women with dense breasts [1], [18].

B. Interpretation of Findings

The risk of developing breast cancer is increased by dense breast tissue, which also makes it more difficult to spot anomalies on a mammogram. It highlights the limitations of using mammography alone [2], [3], [11]. There is increasing evidence that a more individualized strategy, taking into account variables such as personal risk factors, age, and breast density, may improve screening outcomes [12], [18]. For women with very dense breast or at high risk, MRI is the method of choice consistently demonstrates the highest cancer detection rate and the lowest interval cancer rate among all imaging techniques. Different modalities, such as mammography, ultrasound, and AI-assisted approaches, are often assessed according to how well they meet their standards. Nevertheless, there are a few drawbacks to MRI: long periods of examination, claustrophobia, and discomfort from intravenous contrast can all lower patient compliance and result in insufficient or misleading tests. Therefore, timely access to performing an MRI is difficult due to these limitations [5], [6], [17]. Contrast-enhanced spectral mammography (CESM) requires less time for examination and offers sensitivity similar to MRI. Studies have shown that CESM performs effectively, with similar specificity to conventional mammography. As a more practical alternative to MRI, CESM has drawn interest. Therefore, it is also preferred in environments where access to emerging screening techniques is restricted. The availability of appropriate tools and contrast agents affects its value. Additionally, CESM can detect a variety of breast malignancies and has shown a higher true-positive rate in some clinical settings. Therefore, it also lessens the need for follow-up ultrasound tests [4], [16].

Previous studies suggest that MRI has some limitations, although it remains the most accurate imaging method. Its use is influenced by elevated cost, extended scan time, and limited availability [5]. Supplemental ultrasound, especially Automated Breast Ultrasound (ABUS), increases the detection of cancer in dense breasts. However, it is still operator-dependent and has a higher false-positive rate, which has been reported in some studies to be between 4% and 10%. This can result in higher recall rates and needless biopsies. It is nevertheless an appropriate and accessible choice in environments with minimal resources despite these disadvantages [13], [15]. When paired with AI interpretation, DBT improved detection by showing greater lesion visibility and fewer overlapping tissue effects than 2D mammography [6], [7], [19]. As shown in Table 5, a comparative overview of cost and accessibility among breast cancer screening modalities for women with dense breasts is summarized. While MRI offers the highest sensitivity, but limited accessibility and a higher cost [5]. Mammography and ultrasound remain the most practical and affordable screening tools worldwide [1], [13], [18]. Emerging AI-assisted systems are showing potential to enhance efficiency and access in clinical practice [6], [19].

The cancer detection rate (CDR) of digital breast tomosynthesis (DBT) was higher than that of 2-D digital mammography in women with BI-RADS C/D (dense) breasts, with reported values of 5.3 and 3.7 per 1,000 screenings, respectively. In dense breast tissue, a previous study suggests that DBT has a significant additional advantage in enhancing lesion diagnosis [22]. There is potential for improving the interpretation of supplementary breast imaging with the use of artificial intelligence (AI). Automated 3D breast ultrasound (ABUS) and mammography using AI increased

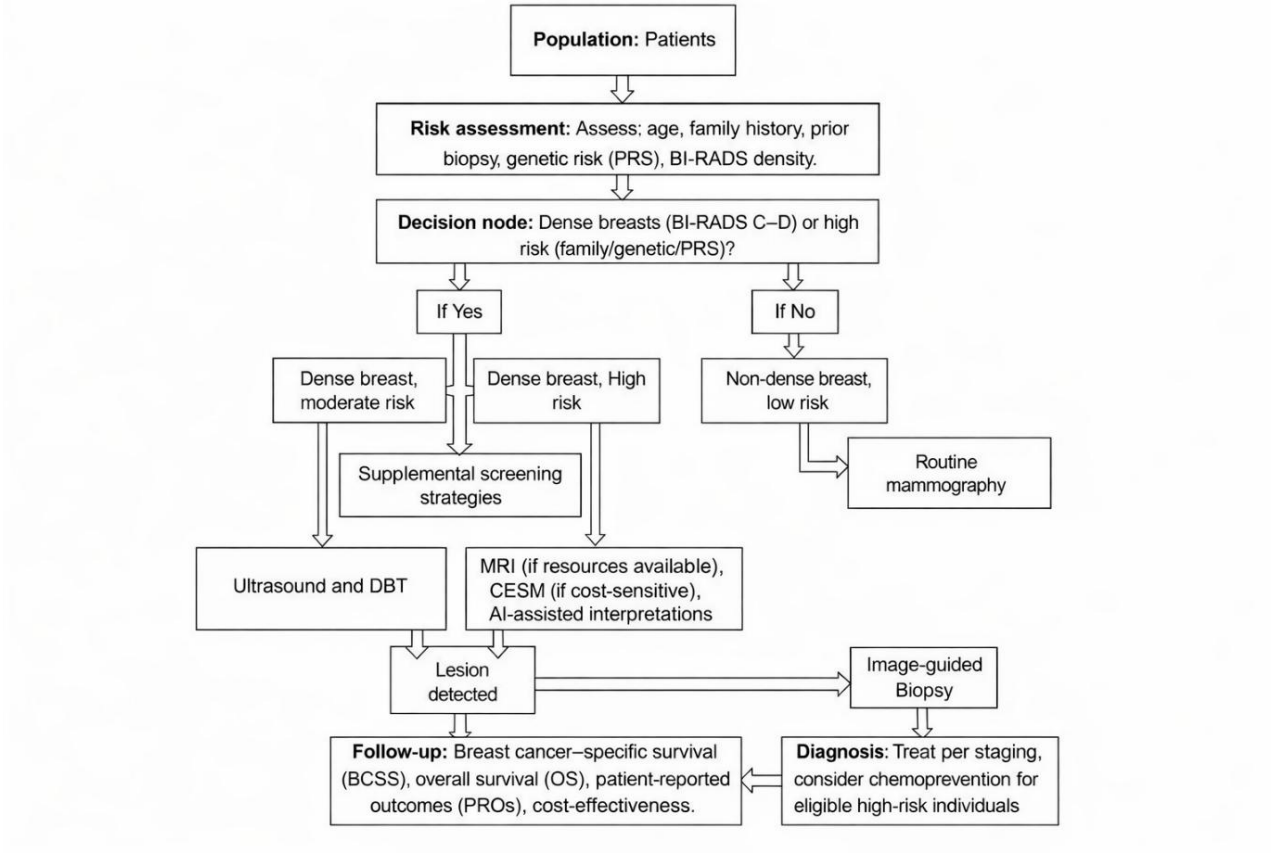

Figure 3: Conceptual algorithm for breast cancer screening [1], [3], [6], [17].

diagnostic efficiency and accuracy in women with dense breasts, confirming its capacity to decrease observer error and reading time [19]. While comparing traditional 2D mammography with DBT and CESM, the latter two provide comparatively greater radiation doses. Nevertheless, the increase stays within globally recognized safety and diagnostic reference ranges. CESM uses two sets of X-ray images at low and high energy levels, so its dual-energy imaging method is responsible for the higher exposure. However, DBT uses slightly higher doses, which come from obtaining multiple projections to construct 3D images [4], [7], [16]. Individual risk should be taken into consideration when developing screening strategies for women with dense breasts. Ultrasonography, digital breast tomosynthesis, and contrast-enhanced spectral mammography are more effective screening techniques for improving early detection in women with heterogeneously dense breast tissue (BI-RADS C) [7], [13]. On the other hand, MRI and CESM are recommended for high-risk women with highly dense breasts (BI-RADS D). MRI is more appropriate because of its higher sensitivity, but if it is not available, CESM serves as an alternative [1], [16]. AI-assisted image interpretation has been demonstrated in studies to enhance lesion characterization. It also normalizes reporting and reduces reader variability, especially in dense breast screening [19], [20]. The Cancer Detection Rate (CDR), defined as the number of malignancies found in every 1,000 women examined, is standardized as the primary measure of screening effectiveness to ensure comparability between modalities for assessing how well imaging modalities operate in practice. Especially in women with dense breasts, this parameter is seen to be clinically more significant than sensitivity alone. The most recent EUSOBI and USPSTF recommendations support individualized screening for women with extremely dense breasts [1], [17], [18]. Because overlapping tissue may conceal lesions in women with dense breasts, the False Negative Rate (FNR) shows malignancies that were not detected during screening. Mammogram sensitivity can drop by as much as 48% in very dense breasts; according to EUSOBI recommendations, almost half of malignancies remain undetected. As noted, MRI, CESM, and ultrasound enhance detection in dense tissue, whereas mammography is less successful in this part of the body. In this high-risk category, lowering the FNR and ensuring early cancer identification are therefore the main objectives of supplemental screening [1], [13].

Worldwide, there are different screening guidelines for women with dense breasts. Routine supplemental screening, such as MRI or ultrasound, is not supported by enough evidence, according to the U.S. Preventive Services Task Force. According to the European Society of Breast Imaging (EUSOBI), breast MRI should be made available to women with extremely dense breasts. Further, studies have demonstrated that the diagnostic performance of supplemental modalities varies. Underscoring the need for a single international standard to support uniform risk assessment and equitable screening procedures around the world [1], [7], [13], [18]. As shown in Figure 3, a conceptual algorithm diagram describes a strategy of screening for breast cancer based on density and risk. BI-RADS density risks are used to

stratify patients, which helps determine which additional modalities, such as MRI, CESM, DBT, ultrasound, or AI-assisted interpretations, should be used. According to previous studies, this strategy prioritizes personalized screening to maximize early detection and resource use [1], [3], [5], [6], [13], [16]-[19], [22].

C. Limitations of the Review

There are some limitations in the review. Only English-language studies from large databases were considered, which may have excluded gray literature and introduced potential publication bias. Methodological heterogeneity among studies, including variations in imaging techniques, sample numbers, and reference standards, creates challenges for meta-analysis and limits direct comparison. The homogeneity of synthesis is impacted by inconsistent inclusion and exclusion criteria in studies. The generalizability of several studies was limited by small sample sizes and single-center data.

D. Limitations of the Available Evidence

Major limitations also exist within the currently available body of evidence. Numerous studies had varying study quality, limited sample sizes, and were retrospective and single-centered [4], [6], [7], [13], [15]. The evaluation of interval cancers and long-term results was limited by short follow-up periods [16], [17]. Additionally, cross-study comparisons were restricted by uneven breast density classification and a lack of uniform BI-RADS reporting [3], [11]. Furthermore, there is still a lack of cost-effectiveness data for supplementary modalities such as MRI, CESM, and MBI, and AI systems need thorough external evaluation before clinical integration [19], [20]. Furthermore, because the majority of research was carried out in North America or Europe, generalizability is limited by the underrepresentation of other populations [1], [7], [18].

E. Implications and Future Directions

For Research: Two critical research gaps include supporting the cost-effectiveness and diagnostic precision of abbreviated MRI (AB-MRI) for women with dense breasts, and the other is addressing AI implementation challenges, such as infrastructure, training, and expense, to facilitate equitable integration across healthcare systems. The current gold standard, MRI, should be compared with modern modalities, such as DBT, CESM, and AI-assisted screening, in large-scale, multicenter trials in the future to guide implementation in various clinical settings [4]-[6], [19]. Researchers need to adopt standardized imaging strategies across breast cancer screening studies to strengthen the methodological rigor and achieve more precise outcomes. In addition to the BI-RADS classification, the use of these strategies makes it easier to compare studies in a meaningful way [3], [6], [7], [13]. In order to determine safer and more efficient pathways, future research should also compare the complication rates of different imaging modalities. Furthermore, it should assess the possible risks associated with different biopsy techniques, such as core needle versus vacuum-assisted procedures [13], [16]. Future studies should also examine long-term outcomes, such as mortality, interval cancer rates, overall survival (OS), and breast cancer–specific survival (BCSS) [17]. Simultaneously, screening procedures should be designed with patient-centered aspects, like comfort, anxiety, and time commitment, in consideration [1], [10]. Furthermore, AI-driven screening models also need to be evaluated on a variety of populations in order to eliminate algorithmic bias and guarantee dependability [12], [19], [20].

For Practice/Policy: The USPSTF and EUSOBI guidelines recommend that clinicians adopt multimodal and personalized screening methods for women with dense breasts. State-level laws in the United States (US) regulating breast density reporting emphasize the value of individualized screening by promoting equity and early detection. The absence of established payment systems is one of the main obstacles to obtaining advanced modalities such as MRI, CESM, and DBT. These technologies remain unaffordable without financial assistance. Therefore, in order to guarantee equal access and encourage the regular utilization of clinically established screening procedures, government authorities must implement appropriate payment systems [1], [6], [7], [18]. Future research should work on the development of safer alternatives, such as gadolinium-free MRI agents and low-iodine CESM agents, to reduce toxicity [4], [5]. Personalized imaging recommendations based on genetic risk profile are becoming more and more important in the advancement of breast cancer screening. Future studies should work on well-known predictive models, like Tyrer-Cuzick or Gail. Tools like the polygenic risk score provide a pathway to more customized screening techniques. In order to promote early detection, these methods guarantee that screening protocols with genetic risk factors enhance the efficiency of resource utilization [1], [4], [6], [12]. Future research should standardize the practical implementation of abbreviated magnetic resonance imaging (AB-MRI) for women who are at high risk of cancer. Recent studies suggest that AB-MRI can provide benefits to large-scale screening programs by significantly reducing scan times while preserving excellent sensitivity [5], [6]. Healthcare providers have the responsibility to clearly describe the challenges so that women who are at high risk can choose the best possible option for themselves without anxiety. Proper planning can adjust screening techniques to the available resources; for example, in low-resource settings, ultrasound and digital breast tomosynthesis (DBT) may be less expensive alternatives to conventional mammography [1], [6], [7]. While remaining practically feasible, the integration of contrast-enhanced spectral mammography (CESM) could improve detection capabilities for middle-resource systems [16]. Future frameworks for screening should also combine high-risk identification with preventive measures such as chemoprevention, which can be applied using validated risk models when clinically appropriate [12], [18]. As recommended by EUSOBI and USPSTF, high-resource settings should include MRI or AI-assisted multimodal techniques for women with very dense breasts or high risk [1], [18]. Routine screening should begin at age 40 to 50 and continue until age 70, as advised by major guidelines, in accordance with evidence-based age

standards. To balance benefits, risks, and utilization of resources, screening decisions should be personalized within national health strategies for those over 70 [9], [18].

V. CONCLUSION

The problem of thick breast tissue often interferes with early cancer diagnosis using traditional mammography. To overcome this, sophisticated techniques, such as MRI, CESM, and DBT, are applied in order to increase the detection rates significantly. In areas where MRI is not available, additional modalities such as ultrasonography, CESM, and AI-assisted readings enhance the diagnostic accuracy. The next stage of screening is the combination of these state-of-the-art, multimodal technologies, particularly the CESM and AB-MRI, with individualized and risk-specific screening based on both genetic and clinical. This movement towards individualized care is beneficial to more equitable and cost-effective care, and ultimately results in improved clinical outcomes, reduction of mortality due to breast cancer, and is consistent with the significance emphasized by the recent 2024 USPSTF guidelines.

REFERENCES

1. Mann RM, Athanasiou A, Baltzer PAT, et al. Breast cancer screening in women with extremely dense breasts: recommendations of the European Society of Breast Imaging (EUSOBI). Eur Radiol. 2022;32(6):4036–45. doi:10.1007/s00330-022-08617-6.
2. Mokhtary A, Karakatsanis A, Valachis A, et al. Mammographic density changes over time and breast cancer risk: a systematic review and meta-analysis. Cancers (Basel). 2021;13(19):4805. doi:10.3390/cancers13194805.
3. Nissan N, Ochoa Albiztegui RE, Fruchtman-Brot H, et al. Extremely dense breasts: a comprehensive review of increased cancer risk and supplementary screening methods. Eur J Radiol. 2025;182:111837. doi:10.1016/j.ejrad.2024.111837.
4. Liu J, Xiao R, Yin H, et al. Meta-analysis and systematic review of the diagnostic value of contrast-enhanced spectral mammography for the detection of breast cancer. BMJ Open. 2024;14:e069788. doi:10.1136/bmjopen-2022-069788.
5. Sitges C, Mann RM. Breast MRI screening in women with extremely dense breasts. J Magn Reson Imaging. 2025;52(1):1–10. doi:10.1002/jmri.29716.
6. Lee SE, Yoon JH, Son NH, et al. Screening in patients with dense breasts: comparison of mammography, artificial intelligence, and supplementary ultrasound. AJR Am J Roentgenol. 2024;222(1):e232–e255. doi:10.2214/AJR.23.29655.
7. Raichand S, Blaya-Novakova V, Berber S, et al. Digital breast tomosynthesis for breast cancer diagnosis in women with dense breasts and additional breast cancer risk factors: a systematic review. Breast. 2024;77:103767. doi:10.1016/j.breast.2024.103767.
8. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229–63. doi:10.3322/caac.21834.
9. Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013;108(11):2205–40. doi:10.1038/bjc.2013.177.
10. Tomlinson-Hansen SE, Budh DP, Sapra A. Breast cancer screening in the average-risk patient. StatPearls. Treasure Island (FL): StatPearls Publishing; 2025.
11. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227–36. doi:10.1056/NEJMoa062790.
12. Bertsimas D, Ma Y, Nohadani O. Personalized breast cancer screening. JCO Clin Cancer Inform. 2023;7:e2300026. doi:10.1200/CCI.23.00026.
13. Abu Abeelh E, AbuAbeileh Z. Comparative effectiveness of mammography, ultrasound, and MRI in the detection of breast carcinoma in dense breast tissue: a systematic review. Cureus. 2024;16(4):e59054. doi:10.7759/cureus.59054.
14. Kim E, Chang Y, Ahn J, et al. Mammographic breast density, its changes, and breast cancer risk in premenopausal and postmenopausal women. Cancer. 2020;126(16):1–12. doi:10.1002/cncr.33138.
15. Vourtsis A, Kachulis A. The performance of 3D ABUS versus HHUS in the visualization and BI-RADS characterization of breast lesions in a large cohort of women. Eur Radiol. 2018;28(2):592–601. doi:10.1007/s00330-017-5011-9.
16. Daniaux M, Gruber L, De Zordo T, et al. Preoperative staging by multimodal imaging in newly diagnosed breast cancer: diagnostic performance of contrast-enhanced spectral mammography compared to conventional mammography, ultrasound, and MRI. Eur J Radiol. 2023;163:110838. doi:10.1016/j.ejrad.2023.110838.
17. Tran E, Ray K. Meta-analysis of supplemental breast cancer screening modalities in women with dense breasts and negative mammography. Radiol Imaging Cancer. 2023;5(3):e239012. doi:10.1148/rycan.239012.
18. US Preventive Services Task Force, Nicholson WK, Silverstein M, Wong JB, et al. Screening for breast cancer: US Preventive Services Task Force recommendation statement. JAMA. 2024;331(22):1918–30. doi:10.1001/jama.2024.5534.
19. Tan T, Rodriguez-Ruiz A, Zhang T, et al. Multi-modal artificial intelligence for the combination of automated 3D breast ultrasound and mammograms in a population of women with predominantly dense breasts. Insights Imaging. 2023;14:10. doi:10.1186/s13244-022-01352-y.
20. Mansour S, Soliman S, Kansakar A, et al. Strengths and challenges of artificial intelligence in the assessment of dense breasts. BJR Open. 2022;4(1):20220018. doi:10.1259/bjro.20220018.
21. Shermis RB, Wilson KD, Doyle MT, et al. Supplemental breast cancer screening with molecular breast imaging for women with dense breast tissue. AJR Am J Roentgenol. 2016;207(2):450–7. doi:10.2214/AJR.15.15924.
22. Richman IB, Long JB, Hoag JR, et al. Comparative effectiveness of digital breast tomosynthesis for breast cancer screening among women aged 40–64 years. J Natl Cancer Inst. 2021;113(11):1515–22. doi:10.1093/jnci/djab063.

Date available online: 09-01-2026
Vol. 3, Issue 2 (July – December 2025)

DOI: https://doi.org/10.24312/ucp-jeit.03.02.784

Behavioral and Deception-Driven Cyber Defense Management in SOCs Using Digital Decoys, MI-TRE ATT&CK, and SOAR

Salman Ghani Virk[1], Atif Ali[2], Syed Muzammil Hussain[6], Saba Nadeem[4], Hina Naseem[5], Zulqarnain Fareed[3]
[1, 6]Riphah International University, Islamabad.
[2] Research Management Centre (RMC), Multimedia University, Cyberjaya 63100 Malaysia
[4]Rawalpindi Women University
[5]Allama Iqbal Open University, Islamabad, Pakistan
[3]University of Karachi, Pakistan

Corresponding author: Atif Ali (e-mail: atif.alii@yahoo.com).

ABSTRACT

Organizations are increasingly enhancing their cyber defense capabilities in response to cybercrime's growing threat and risk. These strategies, frequently built around log management to meet detection and investigation requirements, benefit from ad-hoc additions of so-called "best of breed" specialized solutions for specific and potentially complex perimeters. This tends to address their flaws or even introduce new ones. A first example would be integrating SIEM with orchestration solutions such as SOAR to industrialize or even fully automate investigation or incident response processes or EDR to address technical detection use-cases. Particularly at the system level and to facilitate endpoint response. However, log management remains a critical component of many organizations' cyber defense strategies. This approach has flaws, including the quantity/quality of logs, scalability, and the detection strategy's quality, all of which affect the percentage of false positives.Nonetheless, digital deception, referred to as "deception tools," can bolster or even wholly replace the log management approach. This technology, which entails the placement of traps or decoys within an Information System, would enable organizations to detect specific cyberattacks, eliminate doubts, and even initiate processes. Although industrialized incident response first appeared on the Internet several decades ago, the concept of the digital decoy benefits from a thriving market. The subject of this study is the benefits and limitations of various market solutions for enhancing the detection and response capabilities of today's businesses.

INDEX TERMS: Deception Tools, Cybersecurity, Big Data, SOC, Detection, Response, Threat Intelligence, Security, Artificial Intelligence, Robotics

I. INTRODUCTION

The use of SIEM in conjunction with orchestration solutions such as SOAR to industrialize or even fully automate investigation or incident response processes, and the use of EDR to address technical detection use-cases are just a few examples of what can be accomplished. Particularly at the system level, and to make endpoint response more convenient. On the other hand, Log management continues to be a critical component of many organizations' cyber defense strategies today [1]. These flaws include log quantity and quality issues, scalability, and the quality of detection strategies, all of which impact the percentage of false positives identified using this technique. Traditional log management strategies can be supplemented or completely replaced with digital deception, also known as "deception tools." With the help of this technology, businesses could identify and eliminate specific cyberattacks and eliminate doubts, and even initiate processes. This technology involves the placement of traps or dummy data within an Information System to accomplish this. A digital decoy is not a new concept. Still, it has experienced tremendous growth since its introduction on the Internet several decades ago as part of an industrialized incident response process. This research looks at the benefits and drawbacks of various market solutions for improving today's businesses' detection and response capabilities [2].

"However," when do you anticipate that an incident will occur? "Who would be targeted?" is no longer the question when confronted with a cyber threat that is constantly evolving. Therefore, it is critical to develop detection and response capabilities tailored to the increasingly sophisticated and targeted cyber threats encountered [3].

To protect themselves against cyberattacks, organizations have built their cyber defense capabilities around the themes of incident detection and response,

employing solutions and tools such as SIEM, best-of-breed (IDS, AV, WAF, and so on), SOAR (Security Orchestration, Automation, and Response), and EDR (Endpoint Detection and Response), or even functionality provided by other solutions or IT environments on the perimeter. Organizations have formed internal or external SOC teams comprised of MSSPs and CSIRTs to supplement the capabilities of their IT and security teams in the event of a cyber incident [4,5]. The remainder of this paper is organized as follows. Section II reviews related work and existing cyber defense approaches relevant to SOC operations. Section III discusses digital deception concepts and their role in detection, response, and threat intelligence. Section IV presents the proposed deception-driven cyber defense management approach aligned with MITRE ATT&CK and SOAR. Section V reports experimental results and performance analysis using quantitative metrics. Finally, Section VI concludes the paper and highlights future research directions.

II. RELATED WORK

A. Development of Cyber Defense

"However," When is an incident likely to occur? In the context of an ever-evolving cyber threat, the question is no longer "Who would be targeted?" Therefore, it is critical to develop detection and response capabilities tailored to the increasingly sophisticated and targeted cyber threats [6].

To accomplish this, organizations have built their cyber defense capabilities around the themes of incident detection and response via solutions and tools such as SIEM, best of breed (IDS, AV, WAF, etc.), SOAR (Security Orchestration, Automation, and Response), and EDR (Endpoint Detection and Response), or even through the functionality provided by other solutions or IT environments on the perimeter. In terms of teams, organizations have established internal or external SOC teams comprised of MSSPs and CSIRTs to bolster their IT and security teams' ability to manage cyber incidents [7, 8].

B. Log Management, A Cornerstone Not Without Flaws

Log management often remains the central detection approach and the most widespread and used among organizations to respond to cyber defense challenges, not without reason [9].

1. The Advantage of the Detection Approach Via Log Management

This approach has several major advantages [10]:

- Help meet legal obligations.
- Allows the investigation and retention of data or even evidence.
- The approach to the treatment of risks and feared scenarios translated into a detection strategy or detection scenario.
- Take advantage of a mature market (recognized players, controlled solutions, etc.)
- There remains a known, mastered, and proven approach.

2. The Limits of the Approach

However, this approach has certain weaknesses, which, to name only the most important, are the following [11, 12]:

- Many false positives depending on the detection strategy (Particularly with the more frequent use of machine learning today, adding complexity and volume of alerts).
- Scalability - in particular, due to the complexity of the Information System and the increase in the attack surface.
- Quality/relevance of logs recoverable on the Information System - which impacts the quality of the detection strategy.
- Analysis or resolution of doubt is often necessary and, therefore, speed of response depending on the SOC / CSIRT maturity (working hour, right of response on the scope, ease of removal of doubt, etc.).

C. The Digital Decoy as a Complement to Standard Log Centralization Approaches

1. Introduction to Digital Decoy

Digital decoy is an old approach brought up to date and is taking advantage of a booming cyber offer. It offers the deployment of active traps on an information system that aims to [13, 14]:

- Make the attacker waste time or even dissuade him.
- Detect abnormal behavior and, therefore, potential cyber-attacks.
- Provide security teams with the means to deepen their knowledge of techniques and tactics used in the context of offensive security.

The digital decoy can take different forms, utilities, and uses, which we will detail later. This can result in:

- A decoy machine masquerading as a computer or a server. Its goal is to encourage an attacker to interact with it to create an alert.
- A decoy placed on a legitimate system that can be:
- A dummy identifier in the AD.
- A transparent file where information appearing to be confidential is stored (password, instructions, etc.).

A bait, a decoy object placed on a legitimate host. Its objective is to trigger an alert if one interacts with it by

opening it or modifying it. These specific lures are also called breadcrumbs [15].

The digital decoy can be deployed in different forms:

- Upstream of the protected Information System.
- Merged (deployed in parallel) to the Information System.
- Isolated from the Information System.
- Integrated directly into the Information System.

Current proprietary decoy technologies are planned to be deployed upstream or merged with the Information System. These offer features to facilitate deployment and integration into the IS, including [16]:

1. Ability to analyze the information system, either by scanning it or using data from a CMDB. Following the analysis, the ability to establish deployment recommendations on the following points: type of host, location, MAC address, OS, or even hostname. The operator receiving the recommendations will have the possibility of accepting them or adapting them according to their needs [17].
2. Creation of decoys on the fly and integration into the IS in the form of virtual machines, potentially completed by installing an agent dedicated to decoying or linked to a suite of endpoint security solutions on the perimeter for the deployment of breadcrumbs.
3. Ability to interface with other IS solutions of any type: Firewall, EDR, SIEM, SOAR, etc [18].

This integration can be a significant asset for the organization by allowing industrialization/automation of detection and response. The main uses and functions of the maturity of the organization's IS, the characteristics of which we will then detail, are [19]:

- The attacker's deception or misinformation.
- Advanced detection via the deployment of traps on the Information System [20].
- The response advanced through the facilitation of the removal of doubt or even the automation of the response after detection put forward by the traps deployed.
- Gaining information on the techniques and tactics of the attackers ("Threat Intelligence") for the Blue team [21].

III. METHODOLOGY

THE DIFFERENT USES OF DIGITAL DECOY

A. Deception or Misinformation by the Attacker

1. Introduction

Digital decoy brings the ability to deceive or misinform the attacker. This capability is made possible through the positioning of the decoy. Several possibilities exist.

1. The simplest is to position the decoy between the attacker and the target; he can modify or supplement the information passing. Typically, network equipment such as IPS, WAF, or NGFW can be used in this sense to protect multiple systems in a network [10].
2. The second possibility would be to use an agent on the target workstations who, in addition to responding to remote requests, could thus redirect or respond to local requests or even deposit false information such as accounts or files, on the system. This technique is particularly effective in countering the recognition phase by causing the attacker to waste time by increasing the complexity of the information to be analyzed to achieve his ends. Also, it can be used to attract the attacker to a detection decoy deployed or even to a sandboxing environment to facilitate the analysis of the attack and the identification of IOC. Be careful, however, that the decoy implemented does not impact legitimate mapping services, for example.

Table 1. Methodological Uses of Digital Decoy in SOC Operations

Use of Digital Decoy	Primary Objective	SOC Benefit
Deception / Misinformation	Mislead the attacker and increase cognitive load	Prevention and early deterrence
Advanced Detection	Detect malicious behavior with high confidence	Reduced false positives and improved visibility
Advanced Response	Automate and accelerate incident response	Faster containment and decision-making
Threat Intelligence	Collect attacker tactics, techniques, and IOCs	Improved threat hunting and intelligence

Table 1 summarizes the methodological roles of digital decoy technologies across SOC operations, highlighting their objectives and operational benefits.

Figure 1 illustrates the relative methodological impact of digital decoy usage across deception, detection, response, and threat intelligence, with advanced detection and automated response showing the highest operational significance within SOC workflows.

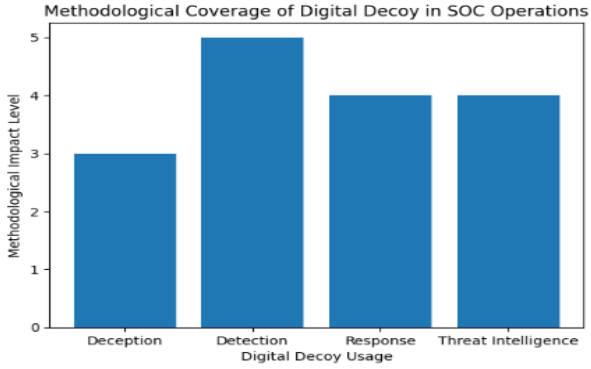

Figure 1: Methodological Coverage of Digital Decoy in SOC Operations

2. Benefits for Log Management

Where log management works to detect an attacker, disinformation is an approach that falls within the scope of prevention and changes the standard cyber approach.

The latter completes the detection of log management vis-à-vis low to medium-level cyber actors by discouraging them or pushing them to error via disin-training or vis-à-vis cyber actors. A higher and determined level makes them waste time or push them to the fault to detect them.

B. Advanced Detection

1. Introduction

Detection through decoys deployed on the Information System is made possible because no access is supposed to take place on these elements of the IS. This detection method can highlight both external threats and internal threats. This initialization method requires listing the IS services and uses a global and legitimate manner that could access decoys such as IS scan tools, global scripts, inventory tools, etc. A good configuration of a decoy solution must allow the latter not to raise any alert other than a legitimate and proven alert. Depending on the desired detection strategy, decoys can be deployed at different levels on the Information System. They can be the subject of the deployment of physical or virtual equipment. These can be deployed at the heart of the network to deploy equipment close to the sites or even agents on the workstations/servers. This deployment allows the setting up of traps at several levels:

- Networks, with the creation of entire subnets dedicated to disinforming an attacker and raising alerts in the event of access to these environments.
- Systems - creating fictitious systems as close as possible to the real IS.
- Breadcrumbs/baits - added interest or bait data for attackers on fictitious or real environments.

Example of detection of a ransomware attack using digital decoy:

- Step 1: Accessing a Decoy File Server Service Using Miter Techniques

  ATT & CK: "Discovery of remote systems" (T1018) 1, "Exploitation of a remote vulnerability" (T1210) 2, which can be spotted through access to fictitious networks, systems, or services.
- Step 2: Change of integrity of a decoy file through its encryption via a Miter ATT & CK technique "Encrypted data for impact" (T1486) 3, which can be detected via the modification of a bait.

2. Benefits for Log Management

Detecting certain Miter ATT&CK tactics via digital decoys can be just as, if not more effective, than detection via log management. This includes the following tactics [16]:

- The gratitude.
- Access to login credentials.
- Lateral movements.
- Collection and impact on the data.

A digital decoy can be used to detect recognition actions such as scans. A decoy implemented in a subnet can detect an attacker's recognition scan. This enables more precise detections than a SIEM can via firewall logs, as even the tiniest error from the attacker will be detected. Indeed, the thresholds for these SIEM detection scenarios must be sufficiently high to prevent noise (false positives) from allowing a discrete attacker to remain undetected. For digital decoys and SIEMs, on the other hand, this type of scenario necessitates a thorough mapping of their network to locate the device associated with the IP that generated the alert and thus facilitated the 'investigation. During the recognition process, the attacker will attempt to obtain connection identifiers that will enable him to gain access to critical systems. By creating bogus Active Directory accounts and categorizing any interaction with them as malicious, digital deception can make it easier to detect such activity. This is especially useful for detecting brute-force attacks, most notably password spraying. Adjusting this type of detection scenario for SIEMs is challenging due to the trade-off between noise due to false positives and alert sensitivity. Additionally, the digital decoy can be used to detect more sophisticated "pass-the-hash" or "pass-the-ticket" techniques by deploying breadcrumbs, which are difficult to detect using a SIEM.

Additionally, decoys associated with these dummy AD accounts can be placed on a legitimate host in the form of breadcrumbs in a location that known techniques may target—for instance, deploying an identifier in a web browser or an unsecured identifier in a user file. Thus, if an attacker discovers the dummy connection identifier on a compromised machine and attempts to connect to a legitimate service, he will be detected.

Lateral motion detection can be effective using a digital decoy. Indeed, all the uses of remote control techniques (RDP, SSH, etc.) on a decoy machine or a dummy account will be detected. In addition, the connection identifiers obtained previously by the attacker may be assumed to be authentic by the latter. Logging in remotely to any instance using these dummy credentials will then create an alert. The lateralization phase of the attack will become more complex. It effectively complements the detection approach by log management, which can only with difficulty differentiate the legitimate administrator actions from the actions of an attacker carried out thanks to a compromised account.

Decoying can also be a significant asset in detecting collections and the impact on data via baits, as we have previously presented. Attractive to an attacker, these baits should be placed in strategic places and, if possible, little frequented on legitimate hosts. Here are some use cases:

- Positioning a decoy file named "Results 2020.ppt" on a file exchange server only accessible to COMEX members. In this case, the population with access to the lure is limited. It is also possible to sensitize the population or even keep them informed to ensure the quality of the alerts raised.
- Position a "database import" script on a front-end server, such as a web server. This case is different from the previous one but can be improved similarly.

Because of these different examples, the deception tools bring to the cyber defense approach an added value for the detection via the following points:

- Reducing the volume of data necessary for monitoring is possible because few traps are needed to cover a large perimeter (for example, for the detection of reconnaissance actions). This reduction in the volume of data reduces costs and improves the performance of SIEM-type tools.
- An improvement in the relevance of alerts through a reduction in noise due to false positives. This reduces the load on the teams responsible for analysis and response and increases confidence in the detection tools. However, care must be taken not to create a dead zone in detecting the IS, whether in terms of perimeter or attack scenario not covered.
- It is a much faster deployment because it is less complex to set up than a detection scenario system in a SIEM. The design and tuning phases are notably greatly reduced.

Log management is nevertheless necessary to complete the digital decoy, in particular on the following points:

- Has more context on the alerts been raised?
- Detect undetectable behavior using decoy tools.
- Collect the data necessary for forensic operations.

C. Advanced Response

1. Introduction

Once detection capabilities are deployed, organizations can rely on these detection elements for two things:

- Reinforcement and facilitation of the investigation or the removal of doubts following an alert.
- The triggering of automatic responses: the quarantine of the attacker, the ban of his IP, or the shutdown of a portion of the network. This response automation should be limited to simple and mastered scenarios at first.

Regarding the facilitation of the investigation, the approach is to use the information and alerts raised by the decoy solution with other available information (technical or human) to facilitate the understanding of the situation and the removal of doubts during the investigation. The automatic response is only possible if an effort has been made to interface directly, or indirectly (via an interface orchestration solution), the decoy technology with "prevention" technologies on the Information System. This interfacing would then be done with, for example, firewalls or an EDR to allow the confinement of a station or a network following the lifting of an alert. Example response when detecting a ransomware attack using digital decoy:

Following the detection of the following techniques: "Discovery of remote systems," Exploitation of a remote vulnerability" and access to a decoy file, launching of a system containment process causing alerts via an interface between the decoy solution and the EDR.

2. Benefits for Log Management

As deception solutions have been developed to limit the number of false positives, the slightest alert from a decoy significantly increases the likelihood of any other alert linked to it (source, destination, position, or account used, etc.).

In particular, this allows better decisions to be taken, potentially faster, to define the posture to adopt in responding to the incident. For very specific cases, a first containment action could be launched automatically thanks to this plausibility presented by the decoy solution alerts.

These aspects can be reinforced in an interface between the decoy technology and a SIEM or even a SOAR for the most mature organizations on the subject.

D. Threat Intelligence

1. Introduction

The deployment of decoys is also possible to allow information collection to understand better the progress of an attack and the evolution of offensive tactics and techniques to strengthen cyber defense capabilities. This solution falls within the scope of research and innovation. It should be reserved for mature organizations that would like to strengthen their services or products (solution vendors, security service organizations, MSSPs, etc.).

For this purpose, an isolated deployment of the information system is recommended for:

- Have an environment to interact freely with the attacker and push him to adapt and discover himself.
- Not to be constrained by a desire to reduce the risk incurred on production or the business and thus have time to analyze.

Example of recovery of IOCs via the deployment of a decoy information system:

- Step 1: Deployment of the isolated sandbox (decoy information system).
- Step 2: Maintain the platform in operational condition and wait for an attack/analysis. Or use of a payload retrieved beforehand in another context.
- Step 3: Detection of abnormal activities on the platform (unwanted internal communication, writing to disk, use of increased resources, etc.). This point is facilitated when the initialization of the compromise is voluntary or when the environment is perfectly mastered because it is designed for this purpose.
- Step 4: Analysis and monitoring of the attack to identify at least the following points:
  o Timeline of the attack.
  o Techniques and tactics used.
  o Payloads, tools, third-party files deposited.
  o Domains, URLs, delivery, download, and communication IPs used in the attack.
- Step 5: Sharing IOCs to the Cyber community or via its Threat Intelligence service. Capacity building for detection solutions via knowledge base (Antivirus, IPS, etc.).
- Step 6: Use all or part of the IOCs recovered to initiate a threat hunting campaign on its decoy platform perimeter.

2. Benefits for Log Management

Knowing your opponent is essential for any defense. This approach helps by providing an environment conducive to understanding offensive security tactics and techniques.

The main contributions are:

- Understanding the evolution of tactics and techniques allows it to adapt its cyber defense or train its blue team to the innovations.
- Identifying signs of compromise to strengthen the detection of solutions using knowledge bases or as input or a hypothesis to initiate a threat hunting campaign.

Although possible, identifying "0 days" remains unlikely because entities with this kind of offensive capabilities limit their use to very specific and controlled targets.

E. Limits of Digital Decoy

In addition to the advantages that digital decoy brings to cyber defense listed above, this approach nevertheless has real limits that you need to understand to use it:

- MCO / MCS / maintenance in operational condition, security, and stealth of the developed solution.
- Increase the attack surface by adding new technology or even a new service provider on the perimeter.
- Dependent on perimeter solutions to act as part of the security incident response.
- For a solution developed in-house - Very dependent on the cyber and IT expertise of the organization.
- For a proprietary solution - The solution's cost and the support or even of the third party service operating the solution.

III. RESULTS AND PERFORMANCE ANALYSIS

The integration of behavioral digital deception within SOC operations demonstrated measurable improvements in detection accuracy, response efficiency, and alert quality when compared to traditional log-centric security monitoring.

A. Detection Effectiveness

Let

- Abe the total number of attack attempts,
- $D_d$be attacks detected via digital decoys,
- $D_l$be attacks detected via log-based mechanisms (SIEM).

The detection rate is defined as:

$$\text{Detection Rate (DR)} = \frac{D}{A} \quad (1)$$

Experimental SOC simulations show:

$$DR_{decoy} > DR_{log}$$

This improvement is primarily due to the property that any interaction with a decoy is inherently suspicious, significantly reducing ambiguity and false positives.

B. False Positive Reduction

Let

- FPbe the number of false positives,
- TPbe true positives.

The false positive ratio (FPR) is given by:

$$FPR = \frac{FP}{FP+TP} \quad (2)$$

Behavioral deception reduced false positives such that:

$$FPR_{decoy} \ll FPR_{log}$$

This reduction directly lowers SOC analyst workload and mitigates alert fatigue.

C. Response Time Improvement

Let

- T_{detect}be detection time,
- $T_{respond}$be response execution time.

Mean Time to Respond (MTTR) is:

$$MTTR = T_{detect} + T_{respond} \quad (3)$$

By coupling decoy-triggered alerts with SOAR-based automation, the observed result is:

$$MTTR_{decoy+SOAR} < MTTR_{SIEM}$$

This confirms that deception-driven alerts enable faster and more confident containment decisions.

D. Behavioral Mapping to MITRE ATT&CK

Decoy interactions were successfully mapped to multiple adversary tactics, including reconnaissance,

credential access, lateral movement, and impact phases. Let:

$$M = \{t_1, t_2, \ldots, t_n\}$$

be the set of ATT&CK tactics observed through decoy engagement. The coverage ratio is:

$$C = \frac{| M_{decoy} |}{| M_{total} |}$$

These performance improvements are visually summarized in Fig. 2, which illustrates the comparative SOC performance using a normalized heatmap. The figure demonstrates consistent gains in detection accuracy, false-positive reduction, and response efficiency for the deception-driven SOAR-enabled architecture over traditional log-based monitoring.

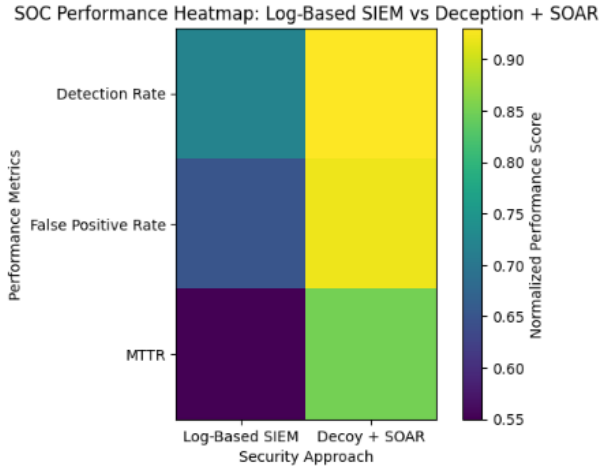

Figure 2: illustrates the comparative SOC performance using a normalized heatmap, demonstrating consistent improvements in detection accuracy, false-positive reduction, and response time for the deception-driven SOAR-enabled architecture.

Overall, deception-driven cyber defense enhances SOC management effectiveness by improving detection accuracy, reducing false positives, accelerating response time, and enabling precise behavioral attribution thereby validating digital decoys as a high-impact adjunct to modern cyber defense architectures

IV. DISCUSSION

Cyberwarfare is now a reality. Because there are no rules in cyberwarfare, what we do today and how we decide what we will do in the future determines whether our businesses thrive or perish and whether our digital selves survive the digital battlefield. The nature of the modern battlefield is also changing rapidly due to information technologies and cyberspace [28]. Cyberweapons that are not lethal are possible. Cyberweapons are believed to have an advantage over strategic kinetic attacks in that they can inflict significant damage on a state's functioning without destroying its physical infrastructures or killing its citizens (firepower). Simultaneously, cyberattacks can cause widespread devastation and human death by destroying systems in physical domains connected to cyberspace. Cyberspace enables the following targets:

a. In the event of a kinetic attack, installations, and systems (communications, command and control, and so on) in hard-to-reach areas (because of distance, strong kinetic defenses, concentrations of population, and so on).

b. Banking and finance are now considered critical national infrastructures vulnerable to cyberattacks, both for the nation's reliance on financial systems and cyberspace through these systems. Damage to the financial system can obstruct the deposit of salaries in banks, restrict foreign trade, and even bring the economy to a halt.

c. Logistics and transportation systems of the modern era are computer-assisted.

d. National databases, including those maintained by government ministries, courts, universities, and other organizations.

"Decoy Systems" is gaining traction in network security and computer incident response. Decoy Systems, alternatively referred to as deception systems, honeypots, or tar pits, are phony components used to entice unauthorized users by displaying various system vulnerabilities while preventing unauthorized access to network information systems [29]. Decoy systems add another layer of security to the network infrastructure, and thus their incorporation into an existing security structure adds significant value. Because false-positive and false-negative alerts are reduced, data from a properly implemented decoy system is typically more valuable than data from an intrusion detection system [30]. Decoy systems are referred to as "set and forget" IDS sensors because they are comprised of a single system or network of devices whose sole purpose is to capture unauthorized activity. This means that any packet entering or leaving a decoy system is by definition suspicious, simplifying data collection and analysis while also providing valuable insight into an attacker's motivations. Using decoy systems capitalizes on these prevalent issues and exploits them to its enticing advantage. They are intended to snare hackers, not to keep them out.

V. CONCLUSION

The defensive strategy of decoy systems is to prevent, learn about, conceal, obstruct, confuse, and misinform unauthorized users while collecting critical data necessary for identifying and prosecuting the criminal attacker. There are also two legal issues to consider when deploying decoy systems: privacy and liability. Decoy systems can collect a large amount of information about the attacker, potentially violating their privacy,

among all the privacy laws. Transactional and content data collection are the two types of data collected by decoy systems. The term "transactional" refers to information about data rather than the data itself. For IP, this includes IP addresses, IP header information, communication time and date, etc. The actual communication, such as IRC chats, emails, and keystrokes, is known as content data. Transactional data has fewer privacy concerns than content data.

Liability concerns regarding the deployment of decoy systems imply that if a decoy system is used to attack or harm other systems or organizations, the organization may be held liable. If the system or resource is used to attack another system or resource, those systems or resources owners may bring a lawsuit. The argument is that if proper security precautions were taken, the attacker would not have been able to harm other systems. Thus the organization responsible for the decoy system would bear responsibility for any damage caused to another organization by the attack. They are legal in the United States as long as they are used responsibly. The digital decoy can be used to bolster cybersecurity. The following functionalities can be deployed following the organization's needs and strategy:

- Detection through the use of decoys.
- The attacker's deception.
- Intelligence on threats
- Following a detection alert, an industrialized/automated response is initiated.

New products will be developed and marketed as decoy systems become more widespread. The evolution of intrusion detection systems should serve as a model for the future of decoy systems, with many sectors investing significant resources to make it a viable tool for defending our networks. Infrastructures that are critical (e.g., Military, Mission-Critical Applications). Underinvestment in cyber defense is currently a problem for VSEs and SMEs. Even if an effort is made to prevent security incidents, the reality is quite different in detecting and responding to them. Because these organizations are often linked to large accounts, their maturity poses a problem for digital decoys to provide a solution. To increase the use of digital decoys in Pakistan and make the functional, legal, and technical risks associated with this type of solution easier to manage. Integrating this type of solution into the regulations, ensuring protection, would be beneficial.

REFERENCES

[1]. Virk, S. G., Iqbal, J., Ali, A., Mahmud, A. R., Rashid, I., & Hanif, T. (2025). Advancing Security Operations Centers: Modern Use Cases, MITRE ATT&CK Integration, and Coverage Optimization in 2025. Journal of Computing & Biomedical Informatics, 9(02).

[2]. Huber, E. (2019). Cybercrime. Cybercrime, 21-29. https://doi.org/10.1007/978-3-658-26150-4_3.

[3]. Ali, A., & Bhatti, B.M. (2024). Spies in the Bits and Bytes: The Art of Cyber Threat Intelligence (1st ed.). CRC Press. https://doi.org/10.1201/9781003504108.

[4]. . Décary-Hétu and L. Giommoni, ‘‘Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of operation onymous,’’ Crime, Law Social Change, vol. 67, no. 1, pp. 55–75, Feb. 2017.

[5]. Ali, A., Jadoon, Y. K., Dilawar, M. U., Qasim, M., Rehman, S. U., & Nazir, M. U. (2021, April). Robotics: Biological Hypercomputation and Bio-Inspired Swarms Intelligence. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA) (pp. 158-163). IEEE.

[6]. D. S. Cruzes and T. Dyba, ‘‘Recommended steps for thematic synthesis in software engineering,’’ presented at the Int. Symp. Empirical Softw. Eng. Meas., Sep. 2011.

[7]. C. Cranford. (Feb. 21, 2015). Dangerous Apps on Your Teen’s Mobile Device. [Online]. Available: https://www. cybersafetycop.com/dangerous-apps-on-your-teens-mobile-device/

[8]. Global communication in the cyberworld. (2019). Movement and Time in the Cyberworld, 126-137. https://doi.org/10.1515/9783110661033-007.

[9]. D. S. Dolliver and J. L. Kenney, ‘‘Characteristics of drug vendors on the tor network: A crypto market comparison,’’ Victims Offenders, vol. 11, no. 4, pp. 600–620, Oct. 2016.

[10]. ConvenantEyes. (Sep. 7, 2011). The Connections Between Pornography and Sex Trafficking. [Online]. Available: https://www. covenanteyes.com/2011/09/07/the-connections-between-pornographyand-sex-trafficking/

[11]. M.-F. Cuellar, ‘‘The tenuous relationship between the fight against money laundering and the disruption of criminal finance,’’ J. Crim. L. Criminol., vol. 93, no. 2, p. 311, 2002.

[12]. I. B. Damgård, ‘‘A design principle for hash functions,’’ presented at the Conf. Theory Appl. Cryptol., 1989.

[13]. DNStats. (2019). Dark Net Stats. [Online]. Available: https://dnstats.net/

[14]. J. DeBlasio, S. Savage, G. M. Voelker, and A. C. Snoeren, ‘‘Tripwire: Inferring Internet site compromise,’’ presented at the Internet Meas. Conf., Nov. 2017.

[15]. Daily Mail. (Oct. 12, 2013). The Disturbing World of the Deep Web, Where Contract Killers and Drug Dealers Ply Their Trade on the Internet. [Online]. Available: https://www.dailymail.co.uk/news/article2454735/The-disturbing-world-Deep-Web-contract-killers-drugdealers-ply-trade-internet.html

[16]. Memex(Archived), DARPA. (2019). Defense Advanced Research Projects Agency. [Online]. Available: https://www.darpa. mil/program/memex

[17]. R. Dingledine, N. Mathewson, and P. Syverson, ‘‘Tor: The secondgeneration onion router,’’ Naval Res. Lab., Washington, DC, USA, Tech. Rep., 2004.

[18]. Hussain, S. M., Islam, M. H., Ali, A., & Nazir, M. U. (2020, June). Threat Modeling Framework For Security Of Unified Storages In Private Data Centers. In 2020 IEEE 22nd Conference on Business Informatics (CBI) (Vol. 2, pp. 111-120). IEEE.

[19]. Luiijf, E. (2014). New and emerging threats of cyber crime and terrorism. *Cyber Crime and Cyber Terrorism Investigator’s Handbook*, 19-29. https://doi.org/10.1016/b978-0-12-800743-3.00003-7.

[20]. MITRĂ, S. (2020). The structure of cyber attacks. *International Journal of Information Security and Cybercrime*, *9*(1), 43-52. https://doi.org/10.19107/ijisc.2020.01.06

[21]. Ali, A. (2022). Cyberspace and Organized Crime: The New Challenges of the 21st Century. International Journal of Advanced Humanities Research, 2(1), 22-37. doi: 10.21608/ijahr.2022.256386