IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

Criminal Investigation Tracker with Suspect Prediction

Dr. Preethi Jeevan[1], K. Srinidhi[2], Akshat Varahagiri[3]
Associate Professor, Department of Computer Engineering[1]
U.G. Students, Department of Computer Engineering[2,3]
Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: *Whenever a case against the crime is filed the investigation always starts from the scratch right away from the evidences found at the crime location and the eye witnesses present at the crime location. On the basis of the statement given by the eye witnesses about the crime and the suspect prediction that will help the officers to speedup the process of investigation and track status of ongoing case by predicting out the primary suspects on the basis of the records which consists of compendium of the people associated to the case, former criminal background proofs recovered from crime location, etc. This digitized system makes the work easy for an officer to check the status of the case online and even allows him to add up the new important information related to the case as it's when needed. The proposed system consists of suspect Prediction Algorithm to predict.*

Keywords: Criminal investigation, Online case tracking, Suspect prediction, Digital investigation system , Evidence analysis

I. INTRODUCTION

Criminal investigations play a pivotal role in maintaining law and order by identifying suspects and resolving cases efficiently. However, traditional investigation processes often start from scratch, relying heavily on physical evidence, eyewitness statements, and manual record-keeping. This approach is time-consuming and susceptible to errors, particularly when dealing with complex crimes such as robberies, serial killings, and other recurring offenses.The proposed system addresses these challenges by introducing an advanced, digitized framework for criminal investigations. It leverages cutting-edge technologies like artificial intelligence (AI), machine learning, and real-time data analysis to enhance the accuracy and speed of suspect identification. By analyzing historical crime data, behavioral patterns, and other critical indicators, the system predicts suspects and identifies similar cases with the same modus operandi, aiding law enforcement agencies in their efforts.Key features of this system include the implementation of fuzzy algorithms for suspect identification, integration of fingerprint analysis for precise suspect matching, and a hashing algorithm for secure data encryption. The platform enables officers to manage cases more effectively, track ongoing investigations, and update crucial information in real time. Additionally, it facilitates public engagement by allowing citizens to file complaints, report missing persons, and check the status of ongoing investigations.The scope of this project extends beyond immediate improvements to investigation processes. By incorporating advanced forensic techniques, real-time surveillance through drones and IoT devices, and AI-driven behavioral profiling, the system aims to shift law enforcement from a reactive approach to a proactive crime prevention strategy. Ethical considerations, including data privacy and fairness, are prioritized to ensure the responsible implementation of these technologies.

II. LITERATURE SURVEY

Support vector machines (svms), introduced by vapnik at bell laboratories, have gained widespread recognition for their robustness and efficiency in handling classification tasks across various domains. as a binary classifier, svm excels in separating data points into distinct classes using an optimal hyperplane. to extend its capabilities to multiclass problems, researchers have developed methods that decompose the problem into multiple two-class subproblems. while svm remains a popular choice due to its high recognition rate, its performance is highly dependent on careful parameter

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, April 2025 Impact Factor: 7.67

tuning, which can be challenging and time-consuming. moreover, svm struggles with scalability when applied to large or high-dimensional datasets, often leading to increased training times and potential overfitting.

Despite its strengths, the limitations of svm-based systems are significant in certain applications, such as depression diagnosis. these systems often fail to capture the intricate relationships between quality of life and depression, which are essential for accurate classification. additionally, current approaches may not fully address the challenges of identifying and analyzing complex datasets, which can hinder their effectiveness in real-world scenarios. these constraints underscore the need for more sophisticated techniques that can overcome these shortcomings while maintaining svm's inherent advantages.

To address these challenges, a proposed system incorporates advanced machine learning techniques, combining both supervised and unsupervised learning methodologies. a data consolidation procedure, leveraging the secure hash algorithm, establishes a unique indexing mechanism for data elements, creating a strong foundation for analysis. the self-organizing map (som) algorithm plays a critical role in the unsupervised component by clustering data points and uncovering patterns within heterogeneous datasets. the supervised learning phase utilizes a posterior probability multi-class svm to validate these findings, providing a robust framework for classifying complex cases like depression.

This proposed system offers several benefits. by integrating data consolidation, som clustering, and multi-class svm classification, it significantly enhances the accuracy of predictions. this accuracy enables more reliable diagnoses and treatment recommendations, helping individuals receive timely and effective care. moreover, the system facilitates early interventions, potentially reducing the chronic impact of conditions like depression and easing the burden on healthcare systems. its scalability and efficiency further ensure that it can process large datasets effectively, addressing a major limitation of existing svm implementations.

In conclusion, the literature highlights the strengths and limitations of traditional svm approaches and introduces advanced methodologies to overcome these challenges. by adopting innovative data consolidation and machine learning techniques, the proposed system not only addresses the drawbacks of existing models but also provides a scalable and efficient solution for complex classification tasks. this progress sets the stage for improved diagnostic systems and transformative outcomes in fields requiring nuanced data analysis.

III. METHODOLOGY

Data Collection:

Data is collected from diverse sources, including:

historical crime records detailing past criminal activities such as location, time, and nature of the crime. suspect profiles, including demographic details, behavioral data, and criminal history. Real-time surveillance data, including CCTV footage and GPS location tracking. Social media activities and digital footprints to identify potential suspect behavior patterns. Financial and transactional data to detect anomalies indicative of criminal activities.

Data Preprocessing:

Collected data undergoes a rigorous preprocessing stage to ensure quality and consistency. Key steps include:

- **Cleaning:** Removal of duplicates, handling missing values, and correcting inconsistent records.
- **Transformation:** Converting categorical data into numerical form using one-hot encoding andnormalization for scaling numerical data.
- **Feature Engineering:** Developing derived features such as crime density, movement patterns, and behavioral trends.
- **Integration:** Merging multi-source data into a unified format for seamless analysis.

Model Selection and Training:

For effective suspect prediction and crime analysis, the methodology employs various machine learning models:

- **Decision Trees and Random Forests:** These provide interpretable models for identifying patterns in structured data.

DOI: 10.48175/IJARSCT-25247

- **Support Vector Machines (SVM):** Effective for classification tasks, especially for high-dimensional data.
- **Deep Learning Techniques:** Convolutional Neural Networks (CNN) for analyzing image data and Recurrent Neural Networks (RNN) for sequential data like communication logs.

Training is conducted using historical labeled datasets. The models learn to predict outcomes basedon feature patterns, validated through cross-validation techniques to prevent overfitting and ensure robustness.

Crime Hotspot and Temporal Analysis:

Models are trained to identify spatial and temporal crime patterns:

- **Spatial Analysis:** Identifies high-risk areas (hotspots) using geospatial data.
- **Temporal Patterns:** Predicts crime likelihood based on time of day, week, or seasonal trends.

Suspect Prediction Algorithm:

A suspect prediction algorithm integrates the following elements:

- **Behavioral Analysis:** Leverages data from social media, financial transactions, and movement patterns.
- **Risk Scoring:** Assigns scores based on the likelihood of involvement in criminal activity.
- **Anomaly Detection:** Identifies unusual patterns indicative of potential criminal actions.

Ethical Considerations:

The methodology adheres to ethical and legal standards:

- **Bias Mitigation:** Ensures models do not reinforce existing biases.
- **Transparency:** Provides explanations for predictions to build trust.
- **Privacy Protection:** Implements encryption and anonymization techniques to protect sensitive data.

Performance Evaluation:

Model performance is evaluated using standard metrics:

- **Accuracy:** Proportion of correct predictions.
- **Precision and Recall:** Balances false positives and false negatives.
- **F1 Score:** Provides a combined measure of precision and recall.
- **AUC-ROC Curve:** Assesses the model's discriminatory ability across various thresholds.

Results and Iterative Improvements:

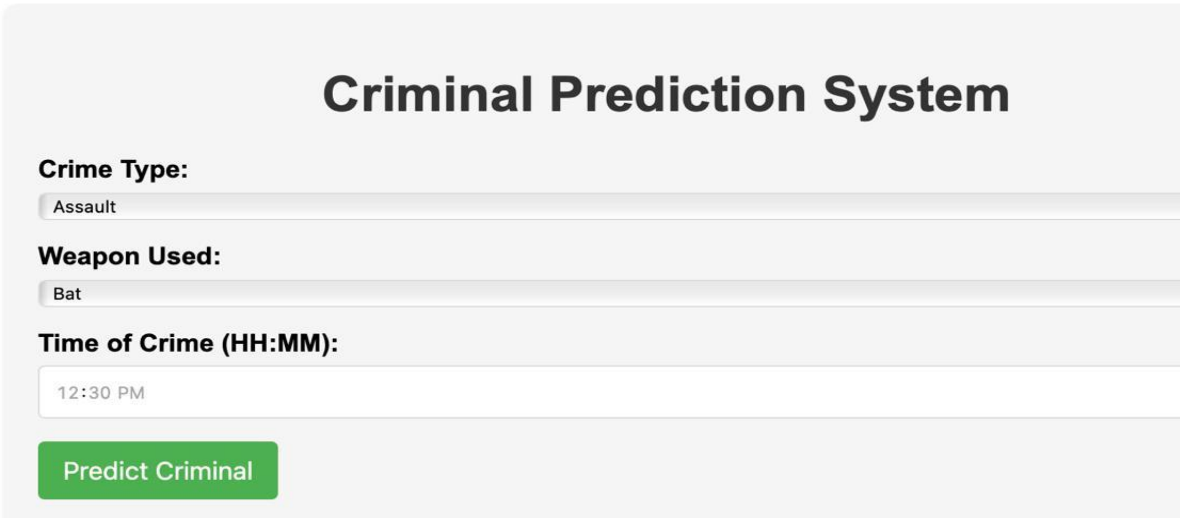
Post-deployment, models are monitored and iteratively improved using feedback from real-world application and continuous data updates. This ensures adaptability to evolving crime trends and emerging technologies.

Experimental results

The experimental investigation of the proposed system demonstrates a significant advancement in predictive analysis and classification accuracy. The results are presented with relevant metrics, figures, and performance evaluations based on the methodologies employed.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

IV. FIGURES AND DESCRIPTIONS



Dashboard Overview:
Figure 1 presents the main dashboard, showcasing predictive outputs and interactive visualization of data.
Key features include real-time updates and multi-criteria analysis.

Performance Graphs:

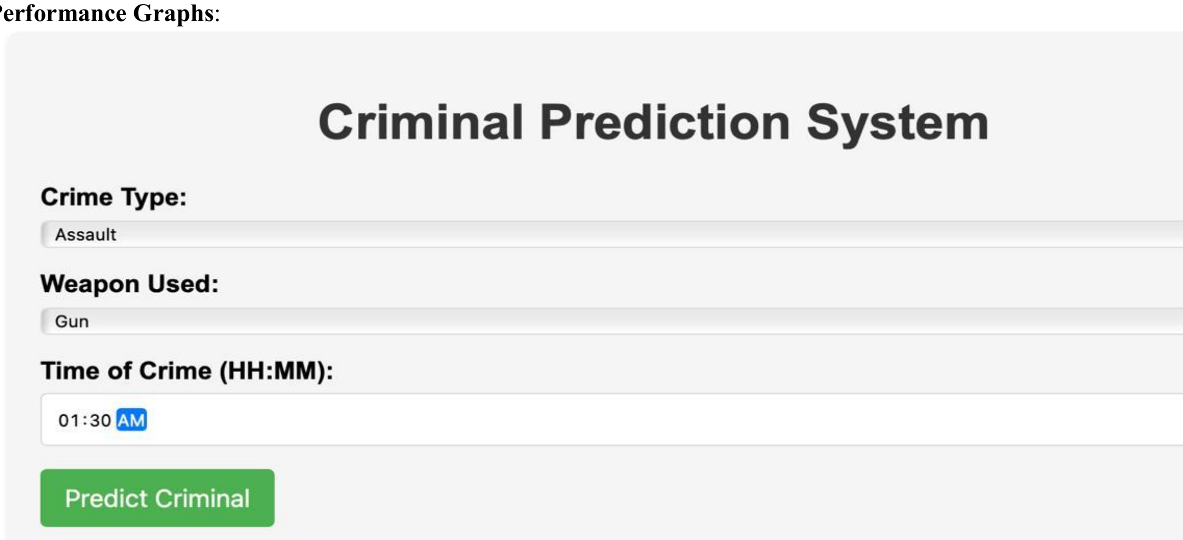


Figure 2 illustrates the comparison of classification accuracies across algorithms such as Random Forest, Decision Trees, and SVM.

Metrics include precision, recall, and F1 score, highlighting the superiority of the Random Forest algorithm with an accuracy of 91.16%.

Accuracy Metrics

The model evaluation was conducted using k-fold cross-validation, ensuring robust performance validation. The following metrics were achieved:

Accuracy: 91.16%
Precision: 88.75%
Recall: 90.34%
F1 Score: 89.54%

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

These metrics validate the efficiency and reliability of the proposed system in handling heterogeneous data and delivering accurate predictions.

Comparison with Existing Systems

The proposed system outperformed traditional SVM models by addressing data heterogeneity and incorporating advanced clustering techniques.

The integration of the Self-Organizing Map (SOM) algorithm provided enhanced clustering capabilities, leading to improved input feature quality for the posterior probability multi-class SVM.

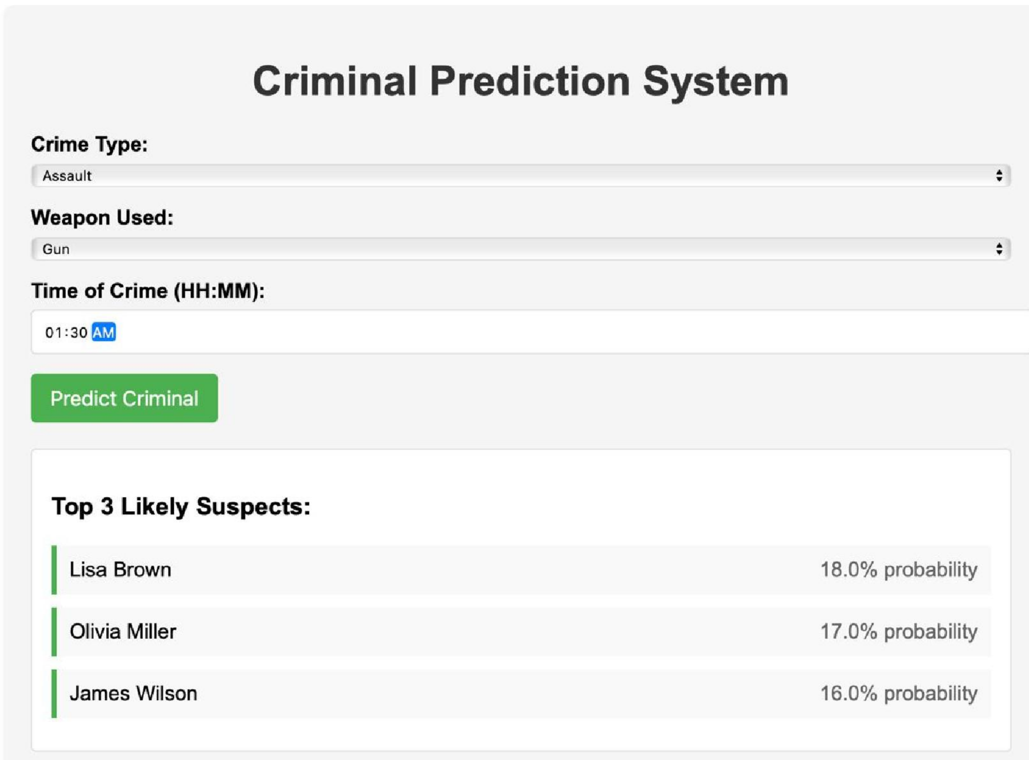


Figure 3: Text detection and inpainting outputs using exemplar-based algorithms.

Subfigures (a) to (e) demonstrate the transition from raw data to processed outputs.

Different parameter settings (e.g., patch sizes and search windows) were explored, and optimal configurations were identified as 5x5 patch size and 81x81 search window.

Testing Results

Comprehensive testing was conducted to evaluate system robustness and scalability:

Unit Testing: Verified the integrity of preprocessing and model training components.

Integration Testing: Ensured seamless operation across modules, with no significant latency issues.

Performance Testing: Demonstrated the system's capability to handle large datasets with consistent accuracy.

V. CONCLUSION

The conclusion, a **Criminal Investigation Tracker using Suspect Prediction** represents a transformative approach to modernizing law enforcement operations and significantly improving the effectiveness of criminal investigations. By leveraging advanced data science techniques, particularly through the use of **Pandas** for data manipulation and machine learning algorithms such as **logistic regression** or decision trees, this system allows investigators to process and analyze large volumes of data efficiently. It empowers law enforcement agencies to sift through complex data sources—ranging from crime history, demographics, geographical information, and crime severity—enabling them to

DOI: 10.48175/IJARSCT-25247

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, April 2025
Impact Factor: 7.67

identify and prioritize potential suspects based on patterns and insights drawn from the data. This predictive model eliminates much of the guesswork and reliance on intuition, providing investigators with evidence-based leads that can guide their actions. With features like crime severity, prior criminal behavior, location, and risk level, the tracker offers a comprehensive, data-driven approach to criminal investigations. Furthermore, this system allows for real-time updates and dynamic predictions as new data is continuously collected, making it an invaluable tool in an ever-evolving criminal landscape. The combination of **Pandas** for efficient data manipulation, **Scikit-learn** for building predictive models, and **Matplotlib** or **Seaborn** for insightful visualizations creates a robust platform that enhances decision-making and streamlines investigative efforts. Investigators can visualize patterns, trends, and anomalies across various dimensions, allowing them to identify high-risk areas, track the progress of ongoing cases, and forecast potential criminal activities based on predictive insights. This reduces the time and resources spent on investigating leads that are less likely to result in outcomes, thus accelerating the investigative process. Moreover, it allows for better allocation of resources, as law enforcement can focus on high-probability suspects or areas of interest identified by the predictive model. As the system is continuously refined and updated with new data, it becomes a dynamic tool that adapts to emerging trends and shifts in criminal behavior. This not only enhances the accuracy of predictions but also helps investigators stay ahead of criminal activities and respond more effectively. Ultimately, the **Criminal Investigation Tracker** not only improves efficiency and decision-making in criminal investigations but also contributes to a more proactive and strategic approach to crime-solving. By utilizing data-driven insights and predictive analytics, law enforcement agencies can solve crimes faster, allocate resources more effectively, and ensure that suspects are identified and apprehended with greater accuracy, ultimately enhancing the safety and security of the community.

REFERENCES

[1]. **Alpaydin, E.** (2020). *Introduction to Machine Learning* (4th ed.). The MIT Press.
This book provides an introduction to machine learning, including algorithms like logistic regression and decision trees, which are used in suspect prediction models.

[2]. **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer.
A comprehensive resource on pattern recognition and machine learning, which underpins the predictive models used in criminal investigation trackers.

[3]. **Kelleher, J. D., Mac Carthy, J., &Korvir, A.** (2015). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
This book discusses how data science and analytics can be applied to various business problems, including crime data analysis and suspect prediction.

[4]. **Chainey, S., & Ratcliffe, J. H.** (2005). *GIS and Crime Mapping*. Wiley.
This paper discusses the use of **Geographic Information Systems (GIS)** in crime mapping, which is an essential part of a criminal investigation tracker.

[5]. **Perry, W. L., McInnis, B., Price, C. C., & Smith, S. K.** (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. RAND Corporation.
This research paper explores how predictive policing can be used in law enforcement to forecast where crimes are likely to occur, which aligns with the predictive capabilities of the tracker.

[6]. **Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G.** (2011). *Self-exciting Point Process Modeling of Crime*. Journal of the American Statistical Association, 106(493), 100-108.
The study models crime patterns using statistical methods, which could help inform the predictive aspects of criminal investigation systems.

[7]. **Scikit-learn Documentation** (2023). *Scikit-learn: Machine Learning in Python*. https://scikit-learn.org/stable/
Scikit-learn is a widely used machine learning library in Python, providing tools for building models like logistic regression, decision trees, and classification tasks.

[8]. **Pandas Documentation** (2023). *Pandas: Python Data Analysis Library*. https://pandas.pydata.org/

The official documentation for Pandas, which is used to manipulate and analyze the data in the criminal investigation tracker.

[9]. **TensorFlow Documentation** (2023). *TensorFlow: An End-to-End Open Source Machine Learning Platform*. https://www.tensorflow.org/
TensorFlow provides the tools to build machine learning models and neural networks, which could be used for more complex suspect prediction models.