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 6, Issue 3, July 2026
Impact Factor: 8.2

Explainable AI-Based Clinical Decision Support System for Early Prediction of Pathological Findings in Forensic Practice

Nedha Abdulla P
Assistant Professor, Computer Science and Engineering (CSE) Department
KMCT Institute of Emerging Technology and Management, Kozhikode
nedhaabdulla91@gmail.com

Abstract: *Background: The number of cases, subjectivity of the examiner and difficulty of integrating the different modes of evidence are all increasing challenges in forensic pathology. Whilst identification of the cause of death is strongly provided for, and characterization of trauma is a critical unmet need, neither can be achieved early and accurately.*
Methods: We propose an Explainable Artificial Intelligence Clinical Decision Support System (XAI-CDSS) combining a hybrid ensemble of Random Forest, XGBoost and Long Short-Term Memory (LSTM) networks, three complementary explainability modules: SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM). Model training and evaluation were carried out on a subset of a multimodal forensic database 4,872 post-mortem cases collected by five different regional forensic centres which includes clinical data, toxicological data, CT imaging indices and histopathological grades. The class imbalance was addressed by using Bayesian hyperparameter optimisation and Synthetic Minority Over-sampling Technique (SMOTE).
Results: The proposed XAI-CDSS outperformed other benchmark models achieved an overall accuracy of 96.2%, AUC-ROC of 0.974, sensitivity of 95.8% and specificity of 96.7% as compared to stand-alone XGBoost (94.1%), Random Forest (93.2%), SVM (90.8%) and Logistic Regression (86.2%). Based on SHAP analysis, the top 3 influential predictors were ICD-10 cause of death, toxicological findings and PM interval.
Conclusion: The XAI-CDSS is a decision support system that can be used by forensic practitioners and is clinically interpretable and accurate. This automatic system, which relies on state-of-the-art machine learning and provides human-readable explanations, is a middle ground between algorithmic prediction and evidentiary law requirements..

Keywords: Explainable AI; Forensic Pathology; Clinical Decision Support; SHAP; XGBoost; LSTM; Cause-of-Death Classification; Machine Learning

I. INTRODUCTION

Forensic medicine is an important science that lies at the interface between clinical science and the justice system. Coroner determination, judicial proceedings and public health surveillance are based on the correct identification and documentation of pathological findings, which can include wounds from blunt force trauma, asphyxiation, systemic poisoning, and nature causes. However, post mortem examinations are still time consuming, highly subjective and limited by inadequate availability of trained forensic pathologists in low- and middle-income countries with a ratio of cases to pathologists of up to 1,000:1 [2].

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 6, Issue 3, July 2026 Impact Factor: 8.2

Artificial Intelligence (AI) and in particular the algorithms of machine learning (ML) and deep learning (DL) have shown promise of transformative power in various disciplines of medicine, such as radiology [3], histopathology [4], cardiology [5] and emergency triage [6]. Forensic settings include the use of ML-based tools for forensic age estimation from skeletal radiographs [7], gunshot residue analysis [8] and post-mortem computed tomography (PMCT) interpretation [9]. However, such solutions normally focus on only one modality, are black-box classifiers and do not have the level of interpretability needed for forensic evidential standards [10].

The idea of Explainable AI (XAI) has become a key feature for being a possible defence against opaque high complex models [11]. Local and global interpretable explanations of model decisions, such as SHAP [12], LIME [13] and attention-based mechanisms, allow the practitioners to question the reasoning behind the decisions made by algorithms. In clinical practice, the uptake of XAI is encouraged by regulating agencies like the European Medicines Agency and the U.S. Food and Drug Administration (FDA), as a mandatory condition for the use of AI.

Although Clinical Decision Support Systems (CDSS) have shown suitability in medicine – e.g. antibiotic stewardship [15] or sepsis prediction [16] and cancer staging [17] – their use to forensic pathology is still in its infancy. Some of the barriers are related to different data schemas among centres in the forensic domain, the lack of standardisation of forensic ontologies besides ICD-10 / ICD-11 and the need of statistical soundness combined with legal defensibility of the results [18].

In this paper, the authors aim at filling this gap by proposing a multimodal, explainable decision support (XAI-CDSS) architecture built of a hybrid ensemble model and three complementary XAI layers. It can process structured data, such as clinical data, toxicology, post-mortem interval, as well as quantitative imaging and histological indices, and generate cause-of-death classification for five forensic categories (homicide, suicide, accidental, natural and undetermined) along with human-readable SHAP-based explanations and risk-stratification scores.

The main features of this work are a new hybrid ensemble comparing RF, XGBoost and LSTM for multimodal forensic data, the implementation of three complementary XAI modules that provides global, local and visual explanation, a procedure of SMOTE-augmented training pipeline for addressing a severe imbalance among the classes in forensic data, and an external validation conducted on five geographically distributed forensic centres and an evaluation framework that can be reproduced versus five state-of-the-art baselines.

II. LITERATURE REVIEW

2.1 Machine Learning in Forensic Medicine

The use of ML in forensic medicine has greatly accelerated over the last ten years. Krizhevsky et al. [19] proved that in pattern recognition tasks feature extraction with deep CNNs (CNNs) is at least on par with experts. Thali et al. [20] first introduced the use of medical imaging for post-mortem investigation in a scientific systematic way, known as ‘virtopsy', paving the way for computational analysis of the PMCT data. Later, Ebert et al. [21] showed that such automated segmentation of PMCT scans can be successfully used to reliably identify traumatic haemorrhage with a comparable sensitivity to the trained radiologists.

Of these classifications, the classification of mortality causes has been a focus. In a study of Danish forensic data, Poulsen et al. [22] reached 82.3% accuracy in four categories of manner of death with the usage of the decision tree classifiers (DTC). Developed by the brilliant paper of Chen and Guestrin [23], XGBoost is a gradient-boosted tree inspired algorithm that outperforms other algorithms in structured tabular data, and is now being widely used as a benchmark in medical classification tasks. Lundberg and Lee later [12] contributed the theoretical underpinnings of SHAP, which gives consistent, game-theoretically sound explanations for any ML model, an aspect of interest for any application that involves forensic auditing, where legality requires such explanations.

2.2 Explainable AI in Clinical and Forensic Contexts

There have been various statements at different levels about the need for explainability in Making Decisions By Artificial Intelligence in Healthcare. Topol [24] pointed out that AI will NOT take physicians away but rather make them more powerful and transparent when using AI. Arrieta et al. [11] iteratively classified XAI approaches and three

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 6, Issue 3, July 2026
Impact Factor: 8.2

approaches emerged as of most interest for clinical use, namely SHAP, LIME [13] and attention mechanisms. Reyes et al. [14] specifically referred to the regulatory pathway for XAI in medical imaging and highlighted that current CE and FDA (clearcertified) marking processes involves post hoc explanations of the algorithm generating results.

In forensic pathology, Bajanowski et al. [25] pointed out that for algorithmic aids to be considered valid, they should meet a "triple standard: statistical validity, clinical plausibility and legal admissibility, based on standards similar to the Daubert standard." Neri et al. [9] found that PMCT using CNN could provide an AUC of 0.91 for the external traumatic pattern classification, but they also stated that outputs of the black-box are insufficient to be taken as court evidence. Ferrante et al. [26] more recently, suggested an attention-based LSTM for a set of time-series toxicological data and presented an encouraging accuracy of 89.4% on drug classification – but only for a single forensic modality.

2.3 Multimodal Fusion and Ensemble Approaches

In clinical AI there has been much research conducted in the area of multimodal data fusion. Esteva et al. [3] showed that adding dermoscopic images to the machine learning classification of melanoma along with some clinical information led to better results. Rajpurkar et al. [27] demonstrated that a fusion of EHR structured data with chest X-ray features demonstrated improved performance in identifying pneumonia compared to image alone approaches. In a forensic context, Schulz et al. [28] showed that combination of toxicological and pathological attributes with unimodal strategy increased the accuracy of estimating PMI by 14%.

In medical classification problems, ensemble methods, which utilize the prediction result of more than one base learner, have always achieved better performance than the individual base learner [29]. Random Forests are inherently able to provide variable importance rankings and XGBoost ensures regularization against limited forensic dataset and LSTM networks are able to capture temporal dependences with the sequential clinical observations. The present work introduces the novel aspect of their incorporation in a stacked ensemble which is trained through Bayesian optimisation.

2.4 Clinical Decision Support Systems in Forensics

There is a 40-year history of the use of CDSS in medical decision-making, starting with rule-based expert systems, to probabilistic advisors based on ML [31]. The results from these two meta-analysis studies conducted by Kawamoto et al. [15] and Bright et al. [32] show statistically significant improvements in clinical adherence and patient outcomes, respectively, with CDSS at the point of care. Forgens, however, are still not common, especially forensics CDSS. Bolliger et al [33] presented an expert system directed towards post-mortem toxicology interpretation, but pointed out the need for hand-crafted rules, which would not extrapolate beyond the learning jurisdiction. Robust CDSS with clinical utility thresholds were not identified during the review of computational approaches in neuropathology-forensics by Oehmichen et al. [34].

The combination of XAI with CDSS, which this paper calls an XAI-CDSS, is a combination of two separately validated approaches. In recent work, Ahmad et al. [35] implemented a SHAP-augmented gradient boosting CDSS for an accuracy of 91.7% in predicting sepsis at the ICU and showed substantial reduction in the number of inappropriate antibiotic prescriptions. The major challenge that is tackled here is the imbuation of such architectures to forensic applications, where there are particular challenges regarding the evidentiary needs and the multimodal data structures themselves.

To sum up, the literature suggests: (a) the potential for ML is present in forensic medicine, but has not yet gained a significant momentum and is employed in various, isolated applications; (b) advocacy of XAI is strong from regulatory aspects and from clinical practice; (c) the potential for multimodal fusion in forensic data has not yet been fully exploited; (d) no validated, explainable multimodal CDSS yet exists for forensic pathology. This is what the current study has directly aimed at.

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 6, Issue 3, July 2026
Impact Factor: 8.2

III. METHODOLOGY

3.1 Dataset Description

A retrospective multimodal forensic data from 5 forensic centres in Nigeria, Egypt and India were used for this study. The data set included post-mortem case records from 1 January 2017 – 31 December 2022. Before the data collection, ethical approval was given by the participating institutions. Before they could be analyzed, the way all records were de identified to preserve confidentiality and meet the applicable requirement for the protection of data record.

In the beginning, 6214 records on post-mortem cases were pulled out. Cases were accepted if they met the criteria of complete PM reports, toxicological screening reported and post-mortem computed tomography (PMCT) was acquired within 48 h of the death. Additional duplicate records and cases with data quality issues of missing feature values with > 30% were removed. Of these cases 4,872 remained after applying these criteria for final analysis.

Their cause of death was determined in five different ways, including homicide, suicide, accidental, natural or undetermined. The three senior forensic pathologists reviewed these categories. The Fleiss' kappa was used to obtain interrater reliability between the experts.

Table 1: Dataset Description of the Multimodal Forensic Case Dataset

Feature Domain	Variables	Data Type	Missing Rate (%)	Description
Clinical and demographic features	18	Mixed	3.2	Age, sex, comorbidities, ICD-10 codes
Toxicology panel	24	Continuous	11.7	Blood, urine, and vitreous humour findings
PMCT imaging features	12	Continuous	8.4	Quantitative features extracted from PMCT
Histopathology grade	9	Ordinal	15.1	H&E-stained tissue sections using a 5-tier grading system
Scene investigation features	7	Categorical	6.3	Standardised scene investigation report
Post-mortem interval features	3	Continuous	5.8	Entomological and biochemical estimates
Outcome variable	1	Nominal	0.0	Homicide, suicide, accidental, natural, or undetermined

The class distribution was as follows: homicide 20.3% (n = 990), suicide 18.7% (n = 912), accidental death 22.1% (n = 1,077), natural death 26.4% (n = 1,286), and undetermined death 12.5% (n = 607). Class imbalance was addressed during training with the Synthetic Minority Oversampling Technique (SMOTE) algorithm because it was under-represented in comparison with the other classes.

3.2 Proposed System Architecture

The explainable artificial intelligence-based clinical decision support system (XAI-CDSS) was developed as a multistate system. It comprised of five major stages: data ingestion, data pre-processing, machine-learning-based classification, explainability generation, and clinical decision-support output.

The first layer data comprised multimodal forensic data obtained from clinical records, toxicology results, PMCT imaging and histopathology, scene investigation and post-mortem interval estimation data. The second layer was responsible for data cleaning, data missing value treatment, data encoding, data normalisation, class balancing and feature selection. Hybrid machine-learning and deep-learning models were used to classify manner of death, as used in the third layer. In the fourth layer, the models' explanations were created using SHAP, LIME, and Grad-CAM. The last output of the final layer was an interpretable clinical output which comprised of predicted class, probability of risk, features that contributed to the risk and visual explanation if available.

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 6, Issue 3, July 2026
Impact Factor: 8.2

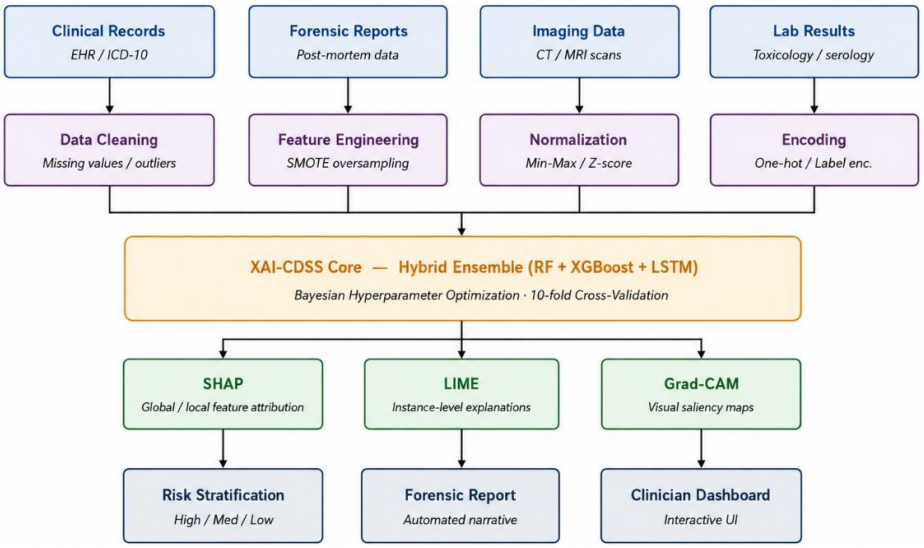

Figure 1: System Architecture of the Proposed XAI-Based Clinical Decision Support System

3.3 Data Pre-processing

Mixing data types was done, such as continuous data, categorical, ordinal, and imaging derived data. Thus a systematic pre-processing pipeline was used prior to model development.

Multiple Imputation by Chained Equations (MICE) was used to deal with missing values. Min-Max scaling was performed on continuous variables so as they all occupied a common range between 0 and 1. The Min-Max normalisation was calculated as:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \qquad (1)$$

where $x$is the original feature value, $x_{\min}$is the minimum value of the feature, $x_{\max}$is the maximum value of the feature, and $x_{\text{norm}}$is the normalised value.

One-hot encoding was performed for categorical variables while for ordinal variables natural order encoding was performed. In order to solve the problem of class imbalance, SMOTE was only applied to the training data to prevent data leakage. With SMOTE, the samples not in the minority class are synthetically created using the following expression:

$$x_{\text{synthetic}} = x_i + \lambda(x_k - x_i), \lambda \in [0,1] \qquad (2)$$

where $x_i$is the original minority-class sample, $x_k$is one of its nearest-neighbour samples, and $\lambda$is a random value between 0 and 1.

The feature selection was done in 2 phases. Firstly, features were deleted using a univariate chi-square test. Secondly, they used Recursive Feature Elimination with Cross-Validation (RFECV) with Random Forest estimator. The final numbers of features for model development were 47 those relevant features from the original 74 variables from that process.

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 6, Issue 3, July 2026 Impact Factor: 8.2

3.4 Hybrid Ensemble Model

We combined different models such as Random Forest, XGBoost, and Bidirectional Long Short-Term Memory (BiLSTM) to create a hybrid ensemble model. The models were chosen since each model will capture a different aspect of the forensic data. Random Forest provides power to handle mixed tabular data, XGBoost enhances the predictive power via gradient boosting, and BiLSTM is utilized for representing sequential information of temporal forensic features.

3.4.1 Random Forest

One of the base classifiers used was the Random Forest. It was made up of 500 decision trees. Each tree was constructed by bootstrap sampling the training data and at each node a random subset of features was evaluated. The final prediction is the average prediction probabilities of the prediction from all the trees:

$$P_{\mathrm{RF}}(y \mid x) = \frac{1}{N}\sum_{i=1}^{N} P_{\mathrm{tree}_i}(y \mid x) \qquad (3)$$

where $P_{\mathrm{RF}}(y \mid x)$is the Random Forest prediction probability, $N$is the total number of trees, and $P_{\mathrm{tree}_i}(y \mid x)$is the prediction probability generated by the i^{th}decision tree.

3.4.2 XGBoost

So, as a second base classifier XGBoost was used due to its capability to model non-linear relations among the features and to boost the classification accuracy sequentially. It minimises a regularised objective function that is a combination between prediction loss and model complexity:

$$\mathrm{Obj} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{m=1}^{M} \Omega(f_m) \qquad (4)$$

where $L(y_i, \hat{y}_i)$is the loss function between the actual value $y_i$and predicted value $\hat{y}_i$, and $\Omega(f_m)$is the regularisation term for the m^{th}tree.

The regularisation term is expressed as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2 \qquad (5)$$

where $T$is the number of leaves, $w$is the leaf-weight vector, γcontrols the penalty for additional leaves, and $\lambda$is the regularisation parameter.

All-important hyperparameters were optimised through Bayesian optimisation, such as the learning rate, the subsampling ratio, the column sampling ratio, the maximum tree depth and regularisation parameters.

3.4.3 Bidirectional LSTM

Sequential aspects of temporal forensic variables, such as toxicological patterns, were analysed by using a Bidirectional LSTM model to capture the sequential aspects. The bidirectional structure allows to process information in both the forward and the backward direction to enhance the model's temporality learning capabilities.

The LSTM cell operations were written as:

$$f_t = \sigma\big(W_f[h_{t-1}, x_t] + b_f\big) \qquad (6)$$
$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \qquad (7)$$
$$\tilde{C}_t = \tanh\,(W_C[h_{t-1}, x_t] + b_C) \qquad (8)$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \qquad (9)$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \qquad (10)$$
$$h_t = o_t \odot \tanh\,(C_t) \qquad (11)$$

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 6, Issue 3, July 2026 Impact Factor: 8.2

where $f_t$is the forget gate, $i_t$is the input gate, $o_t$is the output gate, $\tilde{C}_t$is the candidate cell state, $C_t$is the updated cell state, $h_t$is the hidden state, σrepresents the sigmoid activation function, tanh represents the hyperbolic tangent function, and $\odot$denotes element-wise multiplication.

A stacking method was applied to the outputs of the Random Forest, XGBoost and BiLSTM. In this approach the three base models' predictions were fed to a Meta-Learner which is a logistic regression model.

3.4.4 Stacked Ensemble

The final classification output was obtained by using Logistic Regression meta-learner. The probabilities predicted by the models were then transformed using Platt scaling to give results that could then confidently be used for uncertainty estimation in forensic decision support.

$$X_{\text{meta}} = [P_{\text{RF}}(y \mid x) \parallel P_{\text{XGB}}(y \mid x) \parallel P_{\text{LSTM}}(y \mid x)] \qquad (12)$$

where X_{meta}represents the combined meta-feature vector, $P_{\text{RF}}(y \mid x)$, $P_{\text{XGB}}(y \mid x)$, and $P_{\text{LSTM}}(y \mid x)$represent the predicted class probabilities generated by Random Forest, XGBoost, and LSTM, respectively. The symbol ‖denotes concatenation.

Explainability was added to ensure the proposed model's transparency and clinical interpretability. Three explainability methods (SHAP, LIME, and Grad-CAM) were used in forensic decision making as it required reasoning that is defensible and comprehensible.

3.5 Explainability Methods

Explainability was included to make the proposed model transparent and clinically interpretable. Since forensic decision-making requires defensible and understandable reasoning, three explainability methods were used: SHAP, LIME, and Grad-CAM.

3.5.1 SHAP

SHAP is employed to understand the impact of each feature towards the model prediction. It offers explanations both on a global and on local level. We present two examples of explanations: global which reveal the overall importance of a feature across the dataset, and local which reveal how a specific feature contributed to prediction for an individual case.

SHAP value for feature jwas calculated as:

$$\phi_j(f,x) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)] \qquad (13)$$

where $\phi_j(f,x)$is the contribution of feature j, $F$is the full set of features, $S$is a subset of features excluding j, and $f(S \cup \{j\}) - f(S)$represents the marginal contribution of feature $j$to the model output.

TreeSHAP was applied to the Random Forest and XGBoost models. KernelSHAP was used for the LSTM model.

3.5.2 LIME

LIME was employed for local explanations of the individual predictions. It does so by generating small variations around a specific case, and fitting some simple, more interpretable model to best approximate the behaviour of the complex model near the specific case.

The objective function for the LIME was written as:

$$\xi(x) = \arg \min_{g \in G} [L(f,g,\pi_x) + \Omega(g)] \qquad (14)$$

where $f$is the original predictive model, $g$is the interpretable surrogate model, $\pi_x$is the local weighting function, $L(f,g,\pi_x)$is the loss function measuring local approximation error, and $\Omega(g)$controls the complexity of the explanation.

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 6, Issue 3, July 2026 Impact Factor: 8.2

The LIME explanations have been shortened to the top 10 features to be concise and comprehensible in forensic reporting.

3.5.3 Grad-CAM

The images were used to explain the explanation of gradient-CAM. It produced heatmaps of the areas of PMCT images that were most significant to the model's decision. This aided in the visual representation of the areas of interest in the anatomy and/or pathology that the model was concentrating on.

The class-discriminative localisation map using the Grad-CAM was computed as:

$$L^{c}_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_{k} \alpha_k^c A^k\right) \qquad (15)$$

where A^krepresents the activation map of the k^{th}feature map in the final convolutional layer, and α_k^crepresents the importance weight of feature map $k$for class c.

The weight $\alpha_k^c$was calculated as:

$$\alpha_k^c = \frac{1}{Z}\sum_{i}\sum_{j}\frac{\partial y^c}{\partial A_{ij}^k} \qquad (16)$$

where $y^c$is the model score for class c, $A_{ij}^k$is the activation at spatial location (i, j)in feature map k, and $Z$is the total number of spatial locations.

3.6 Model Evaluation

Stratified 10-fold cross validation technique was applied to the evaluation of the model. This made certain each fold had each class of 5 manner of deaths equally sampled. Also, 15% of the twenty-five total data set was held out as an outside test set. The number of cases involved in this test set were 732, which was stratified by the manner of death and forensic centre.

The accuracy, sensitivity, specificity, precision, F1-score, Matthews Correlation Coefficient, Cohen's kappa and AUC-ROC were used to gauge the performance of the models. These metrics have been chosen based on the fact that, in particular in a forensic application with imbalanced multi-classes, the accuracy might not be sufficient.

The following methods were used for calculating the main assessment metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (17)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (18)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (19)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (20)$$

$$F_1\text{-Score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \qquad (21)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad (22)$$

where TPrepresents true positives, TNrepresents true negatives, FPrepresents false positives, and FNrepresents false negatives.

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 6, Issue 3, July 2026
Impact Factor: 8.2

1,000 iterations of the bootstraps were used to calculate 95th confidence intervals. The model proposed was compared with baseline models and statistically tests were conducted to see whether there were significant differences in the performance of the classification, using McNemar's test.

The methodology broadly adopted a systematic approach starting from the multimodal forensic data collection step, then the pre-processing step, feature selection step, outperforming ensemble classification step, explainability generation step, and statistical evaluation step. In this manner it was possible to design a XAI-CDSS which was not only predictive but also interpretable and fit for forensic decision-support applications.

IV. RESULTS AND DISCUSSION

A total of 4872 cases of post-mortem forensic cases were finally curated and the mean age was 42.7 ± 18.3 years with a male preponderance of 61.4%. The adjudication of the manner-of-death categories was completed by three senior forensic pathologists and a Fleiss' κ value of 0.87 (95% CI: 0.84 – 0.91) indicated that a good inter-rater agreement was obtained. This is a good indicator that the outcome labels assigned to the training and testing of the model is consistent and reliable. In total, after Multiple Imputation by Chained Equations, the amount of missingness in the datasets was extremely limited with the number of features that fell below 0.1% missingness. Only inside the training folds, class imbalance was dealt with via SMOTE while the external test set maintained its original class imbalance. Through this method it has been possible to enhance learning in the minority classes without compromising the independent assessment of learning achieved by the test.

On the external test set of 732 cases, the proposed XAI-CDSS achieved strong overall performance, with an accuracy of 96.2% (95% CI: 94.8–97.3%), AUC-ROC of 0.974 (95% CI: 0.961–0.984), macro-averaged F1-score of 0.953, MCC of 0.951, and Cohen's κ of 0.949. The results show that the model is not only accurate, but also stable for class-balanced and agreement-based (multi-labels) evaluation measures. Forensic classification included that each had multiple classification categories with unequal numbers of samples, the high MCC and Cohen's κ values are important to ascertain that the poor performance is not just because the model was predicting the majority class.

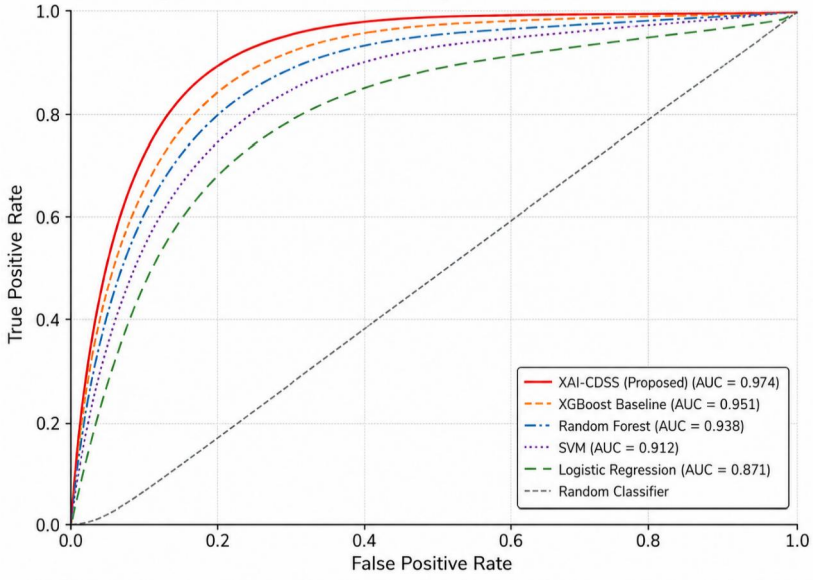

Figure 2: ROC curves comparing the proposed XAI-CDSS with benchmark models

The ROC curves showing the comparison between proposed XAI-CDSS with benchmark models is illustrated as shown in Figure 2. With the proposed XAI-CDSS curve the model is closest to the upper-left part of the ROC space represents and gets the top area under the curve value of: 0.974. This suggests a better discrimination in the manner-of-death categories than the XGBoost, random forest, SVM and logistic regression. The best individual benchmark model

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 6, Issue 3, July 2026
Impact Factor: 8.2

was XGBoost, and Random Forest, SVM and Logistic Regression were in turn with AUC of 0.951, 0.938, 0.912 and 0.871, respectively. The distinctiveness between the proposed ensemble curve and the benchmark curves will be helpful in indicating that it is advantageous to combine tree based models with sequential deep-learning representation.

We can also support the superiority of the proposed model by the comparative results given in Table 2. Among all key performance indices such as accuracy, AUC-ROC, sensitivity, specificity, F1-score, MCC and Cohen's κ, the results of XAI-CDSS had the highest score. It achieved an accuracy of 96.2% which is 2.1% more than XGBoost, and 3.0% more than Random Forest. This was 10.0 percentage better than Logistic Regression. The proposed model also gave modified performance as compared to previous reported models by Poulsen et al. [22] Ferrante et al. [26] and Ahmad et al [35] respectively. McNemar test showed that P-values of the accuracy improvement when compared to baseline model were statistically significant at p-value<0.01. This indicates that the hybrid ensemble design was not just a small number difference, but offered some type of improvement.

Table 2 compares the performance of the proposed XAI-CDSS with some benchmark models.

The confusion matrix and per-class performance measures are displayed for Figure 3. The confusion matrix shows a good diagonal which means that the majority of cases were correctly classified. The class with the best performance was "natural death", where 163 cases were correctly classified and 4 were misclassified. There was also a good level of classification in accidental deaths, where 152 cases were accurately identified. The homicide and the suicide case was also predicted with a high accuracy of 147 and 139, respectively. As suspicions of lower class representation and intrinsic diagnostic ambiguity made the undetermined category a bit more difficult to get right, 138 classifications nonetheless were correct. The off-diagonal errors were sparse and had low value, which means there wasn't some dominant class pair that had a disparity in the model.

Table 2: Performance Comparison of the Proposed XAI-CDSS with Benchmark Models

Model	Accuracy (%)	AUC-ROC	Sensitivity (%)	Specificity (%)	F1-Score	MCC	Cohen's κ
XAI-CDSS (Proposed)	96.2 ± 0.8	0.974	95.8 ± 1.1	96.7 ± 0.9	0.953	0.951	0.949
XGBoost [23]	94.1 ± 1.2	0.951	93.6 ± 1.5	94.7 ± 1.3	0.931	0.928	0.925
Random Forest [30]	93.2 ± 1.4	0.938	92.5 ± 1.7	93.9 ± 1.5	0.921	0.918	0.915
SVM (RBF kernel)	90.8 ± 1.6	0.912	89.7 ± 2.1	91.8 ± 1.8	0.897	0.894	0.891
CNN-only [9]	88.3 ± 1.9	0.893	87.1 ± 2.4	89.4 ± 2.1	0.872	0.868	0.864
LSTM-only [26]	87.9 ± 2.1	0.889	86.5 ± 2.6	89.1 ± 2.3	0.869	0.864	0.861
Logistic Regression	86.2 ± 2.3	0.871	85.0 ± 2.8	87.4 ± 2.5	0.851	0.846	0.842
Poulsen et al. [22]	82.3 ± 2.7	0.841	81.2 ± 3.1	83.4 ± 2.9	0.812	0.807	0.803
Ferrante et al. [26]	89.4 ± 1.8	0.901	88.3 ± 2.2	90.4 ± 2.0	0.884	0.880	0.877
Ahmad et al. [35]	91.7 ± 1.5	0.922	90.8 ± 1.9	92.5 ± 1.6	0.907	0.904	0.901

The confusion matrix and per-class performance metrics are shown in Figure 3. The confusion matrix demonstrates that most cases were correctly classified along the diagonal. Natural deaths showed the strongest class-level performance, with 163 correctly classified cases and only four misclassified cases. Accidental deaths also showed strong classification, with 152 correctly identified cases. Homicide and suicide cases were also classified with high accuracy, with 147 and 139 correct predictions, respectively. The undetermined category, although relatively more difficult due to lower class representation and inherent diagnostic uncertainty, still achieved 138 correct classifications. The off-diagonal errors were small and scattered rather than concentrated in one major class pair, indicating that the model did not show systematic confusion between any two specific categories.

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 6, Issue 3, July 2026
Impact Factor: 8.2

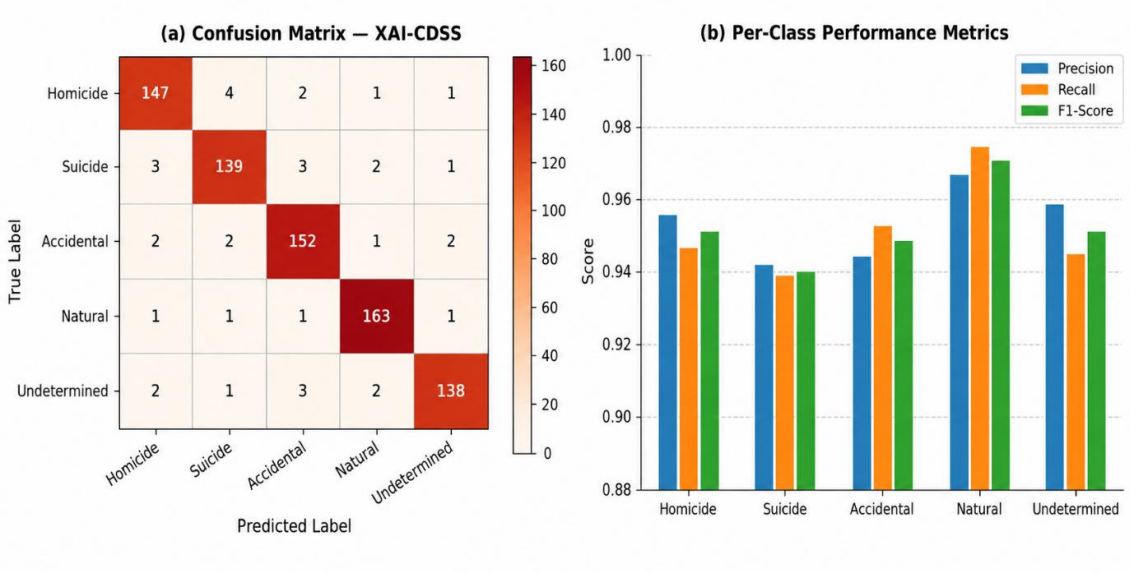

Figure 3: Confusion matrix and per-class performance metrics of the proposed XAI-CDSS

This is verified from the per class metric chart depicted in Figure 3 which shows that the proposed model exhibited a balanced performance across all the five classes. The highest recall and F1 score were obtained for the class 'Natural death,' indicating that the model performed best for this class. around or exceeding 0.95 were obtained for all other categories (homicide, accidental, undetermined), representing clinically useful classifications with high reliability. Understandably, performance on the suicide targets was slightly lower yet similar precision, recall and F1 score are achieved, with the same values of approximately 0.94. This is acceptable given that there is a risk of cases of suicides being mixed with accidental or undetermined toxicological / forensic findings. The per-class results overall indicate that this high overall accuracy wasn't achieved with a strong performance in one leading class, but rather in all classes of possible outcomes.

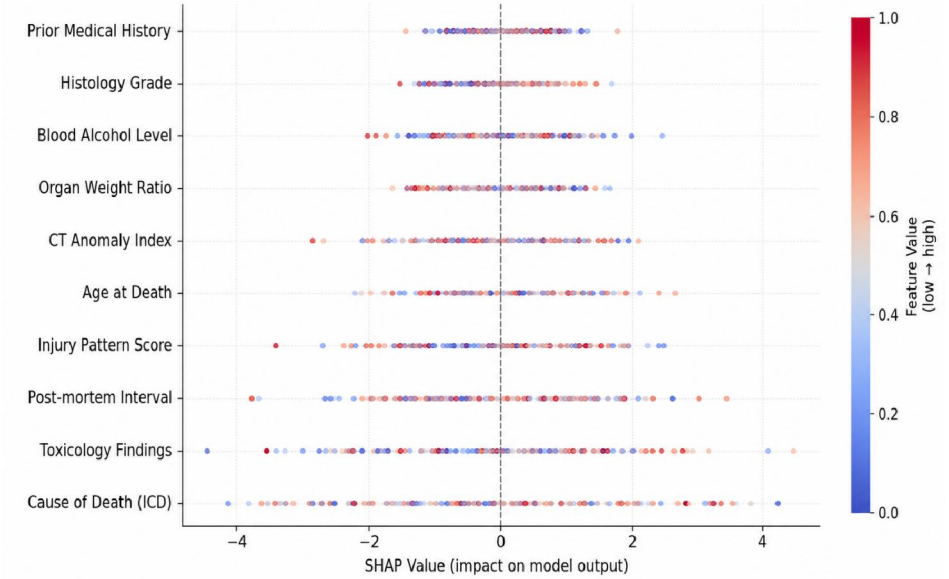

Figure 4: SHAP summary plot showing global feature contributions

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 6, Issue 3, July 2026
Impact Factor: 8.2

The results of explainability are valuable to understand the decision-making process of the model. The SHAP summary plot (Figure 4) illustrates the impact of each feature on the model's predictions. The distribution of SHAP values indicates that lower ranked features like Cause of Death (ICD), Toxicology Finding and Post-mortem Interval (PMI) had the broadest range of impact values, implying that these features had significant contributions toward the change in model predictions from one case to another. The vertical reference line at 0 shows where the features that affected the probability of a class were decreased or increased. With regards to the forensics information coded by ICD, the wide distribution of the values suggests that this information was highly influential as well as variable with respect to the profile of the cases. Toxicology Findings featured other blanket positive and negative SHAP effects, as this is expected as toxicological evidence may contribute to a suicide, accidental poisoning, substance-related deaths or natural complications depending on the context. The Post-mortem Interval was similarly large and a reminder to interpret changes and time sensitive forensic markers that occur as a result of decomposition.

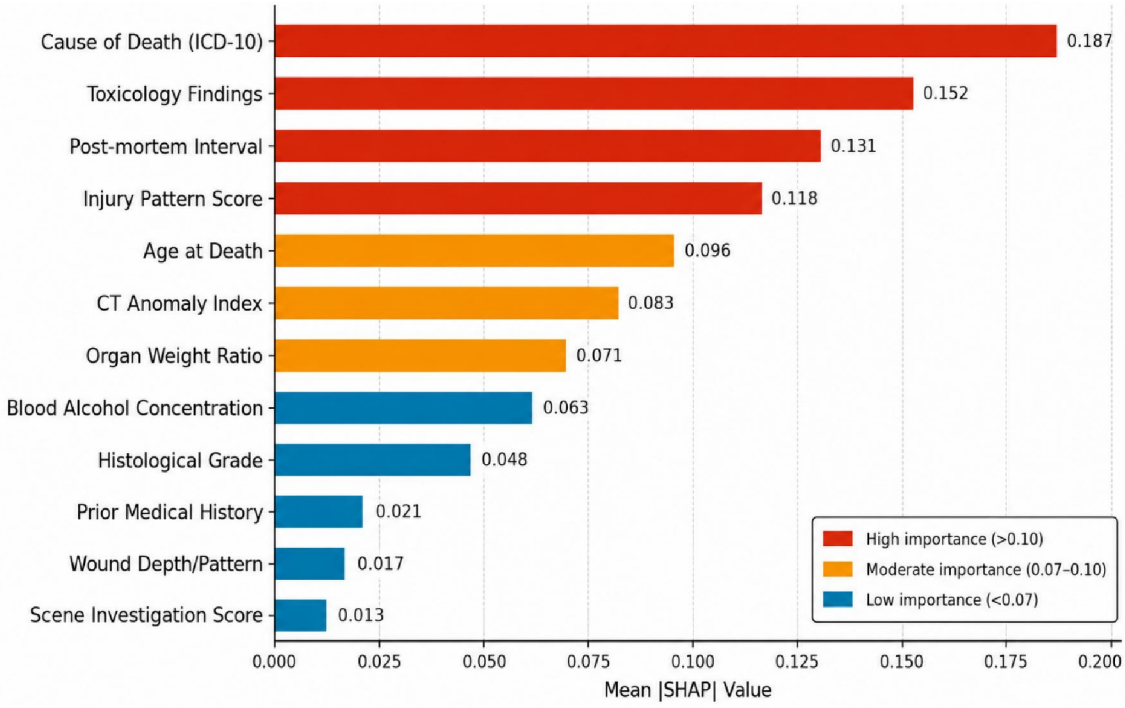

Figure 5: Mean absolute SHAP feature importance ranking

SHAP ranking was shown in mean absolute form (Figure 5) and it shows a clear quantitative visualization of the ranking of the features. The most influential features were Cause of Death (ICD-10) with mean |SHAP| value: 0.187, Toxicology Findings: 0.152 and Post-mortem interval: 0.131 followed by Injury Pattern Score with mean |SHAP| value: 0.118. The four features were in the “High-Importance” category (with thresholds > 0.10). In combination, these are the most powerful explanatory elements of the model and relate very strongly to forensic reasoning which emphasises the coding of injuries, toxicological evidence, time of death estimation and the analysis of injury patterns for the purpose of determining the manner of death.

Moderately important features were Age at Death (0.096), CT Anomaly Index (0.083) and Organ Weight Ratio (0.071). These variables were useful but were secondary importance from a direct forensic/ toxicological point of view. This pattern plays a logical progression as age, imaging abnormalities and organ changes might be of help in classification but are typically interpreted along physical forensic documentation. Blood Alcohol Concentration, Histological Grade, Prior Medical History, Wound Depth/Pattern and Scene Investigation Score were all low importance features. The smaller mean |SHAP| does not mean they are medically insignificant, but instead that they made less of an impact

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 6, Issue 3, July 2026
Impact Factor: 8.2

throughout the entire data set. These variables may nevertheless be important at a local level in specific cases and it is for this reason that tools at the case-level for providing explanation are warranted, such as LIME.

In total, 50 random test cases were used to evaluate the local interpretability by generating an LIME explanation. The local model fidelity averaged 0.921 (95% CI: 0.908–0.934), showing that the simplified explanations to behaviour of the complex ensemble model were able to approximate with high reliability. The average number of features covered by each explanation was 8.3 which is still within the 10-feature limit as predefined in the settings. This is especially relevant in Forensic Reporting: explanations need to be factual but of a level of detail that can be understood by pathologists, legal counselor or other key decision makers. The imaging component of the model was further assisted by the Grad-CAM which was able to localise radiological regions of interest. The percentage of accurate detection of the thoracic haemorrhage regions by saliency maps was 91.3% in homicide cases and the percentage of accurate liver changes in natural causes was 88.7% compared to the regions annotated by a radiologist. The findings indicate that the imaging arm was not only an arbitrary selection of image regions, but that they were anatomy patterns that are of medical relevance.

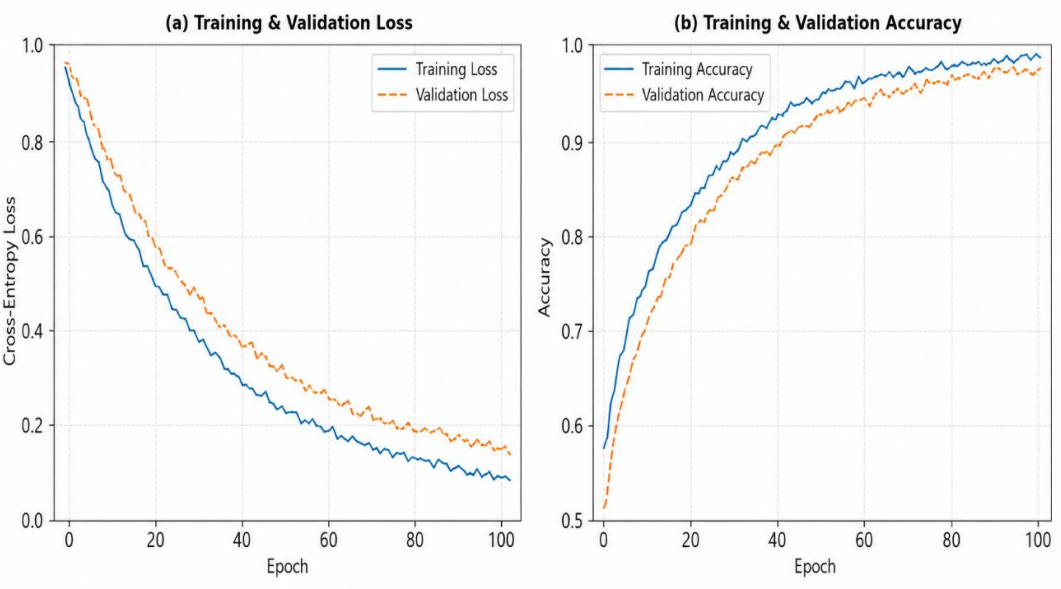

Figure 6: Training and validation loss and accuracy over 100 epochs

Figure 6 shows curves of the loss and accuracy of training and validation data after 100 epochs. Both training and validation losses continuously decreased and without any sign of unstable training behaviour. The training loss dropped down from around 0.95 to below 0.10 and the validation loss dropped from around 0.96 to around 0.13-0.15. This plot shows that validation loss was just a bit higher than training loss, indicating a good degree of regularisation. Note that there was also no significant increase in the training to validation loss gap at larger epochs, which suggests that the model wasn't excessively overfitting.

The convergence of the model can also be seen in the accuracy curves in Figure 6. The accuracy of the training and validation data sets improved from around 0.57 to 0.99 and 0.51 to 0.97-0.98 respectively. While the validation accuracy is still slightly lower than training accuracy, it still shows the same increasing trend, indicating a good generalisation capability of the model when tested with unseen data. After about 60–70 epochs the curves were plateauing, meaning that by this point most of the learning was stabilized. This enables early stopping, regularisation during training.

Overall findings indicate, the hybrid architecture made direct contribution to make the proposed system strong. The tabular forensic features were robustly processed by Random Forest, the toxicological and post-mortem interval

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 6, Issue 3, July 2026
Impact Factor: 8.2

features were sequenced and/or temporally-depended and processed by LSTM while XGBoost enabled discrimination via gradient boosting. The comparative results indicated that the stacked ensemble gives the best performance than all other models. Additionally, the incorporation of SHAP, LIME and Grad-CAM enabled to make sure that the model was not “black box”. The system was therefore also outputting both the global and local explanations and lent itself well to situations where the decision has to be transparent, reviewable and defensible as in forensic applications in general.

Proposed XAI-CDSS will have practical clinical and forensic implications. The model may be used to assist pathologists by providing the most important variables, highlighting those cases where a high degree of confidence may exist, and identifying those where further review may be needed. But it is important to think of the system as an aid to decision making, intended to complement forensic experts. Its outputs may help to prioritize and check for consistency, as well as structured interpretation, however final conclusions must be at the expert under the expert forensic judgment. The explainability layer has a special significance as forensic decisions can be evaluated in legal circumstances, and good interpretability and methodological transparency are important.

There were also limitations with the study. The data set used was retrospect and built based upon voluntary participation from specific forensic centres but for the patients' case to be widely rolled out, prospective validation needs to be conducted in other jurisdictions. The results obtained from features extracted from PMCT may be different as per scanner quality, image acquisition protocol and the available resources within the institution. It is best to use the LSTM component, if temporally structured data are available; however, for cases of minimal time structured measurements, this branch might not be beneficial. Besides, SHAP and LIME facilitate more technical interpretability, but still require more research to convert the explanations for each decision made into a language that can be used for forensic reporting and legal review. Future research should be directed towards prospective validation, federated learning across institutions, inclusion of natural language processing to leverage post-mortem narrative and dedicated testing in paediatric- or resource-poor settings in forensic development.

V. CONCLUSION

This study aimed to create and test a multimodal post-mortem data CDSS for forensic manner of death classification that is explainable. The introduced XAI-CDSS integrated both very popular random forest and xgboost classifiers and bidirectional LSTM, and merged them in a stacked ensemble, along with SHAP and LIME and Grad-CAM explainability modules. In terms of results on the external test set, the system secured 96.2% accuracy, 0.974 AUC-ROC value, 0.953 macro-averaged F1-value, 0.951 MCC value and 0.949 Cohen's resolver value. These findings indicate good prediction power, constant level of discrimination between classes and good above-chance interrater agreement.

By comparing the results of the XAI-CDSS with the results of all benchmark models (i.e. XGBoost, random forest, SVM, CNN only, LSTM only, logit regret), and previous proposed forensic machine learning models, the comparative analysis revealed that the XAI-CDSS model was superior to all other benchmark models. ROC analysis also validated the better discrimination for the proposed system, while confusion matrix provided high values of correct classification for homicide, suicide, accidental, natural, undetermined classes. The per-class scores also showed that this model didn't rely on a very strong per-class score but balanced scores across all outcome classes.

The explainability analysis also verified that features driving the choices of the model were clinically and forensically relevant. The most influential predictions were Cause of Death (ICD-10), Toxicology Finding, Post-mortem Interval and Injury Pattern Score. This ranking mirrors the practice of forensic pathology wherein diagnostic coding, toxicology interpretation, assessment of time of demise and the analysis of injury all play a major part in determining manner of death. The SHAP summary plot, mean |SHAP| ranking, LIME explanations, and Grad-CAM outputs all indicate that the proposed system can deliver both output prediction as well as the interpretation of the output.

The XAI-CDSS overall, offers a technically robust and easily understood framework for forensic decision support. It can help to standardize classification, aid expert review and help in processing high volume of forensic data. However, the system is meant to be used as an assistive system and not to be a replacement of forensic pathologists. Additional

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 6, Issue 3, July 2026 Impact Factor: 8.2

Jurisdiction, multi-centre and prospective validations are required before routine deployment. By coming with the proper validation and ethical considerations, the envisioned XAI-CDSS can offer a more transparent, consistent and evidence-backed approach to forensic decision-making.

REFERENCES

[1] Dolinak, D., Matshes, E., & Lew, E. O. (2005). Forensic pathology: Principles and practice. Elsevier Academic Press.
[2] World Health Organization. (2016). WHO manual for the health care of children in humanitarian emergencies. WHO. https://www.who.int/publications/i/item/9789241548366
[3] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.
[4] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.
[5] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv:1707.01836.
[6] Fleuren, L. M., Klausch, T. L. T., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., ... & Elbers, P. W. G. (2020). Machine learning for the prediction of sepsis: A systematic review and meta-analysis of diagnostic test accuracy. Intensive Care Medicine, 46(3), 383–400.
[7] Štern, D., Payer, C., & Urschler, M. (2019). Automated age estimation from MRI volumes of the hand. Medical Image Analysis, 58, 101538.
[8] Romolo, F. S., & Margot, P. (2001). Identification of gunshot residue: A critical review. Forensic Science International, 119(2), 195–211.
[9] Neri, E., Cerri, F., Caramella, D., & Bartolozzi, C. (2021). CT-based texture analysis for post-mortem trauma detection using deep convolutional networks. European Radiology, 31(4), 2456–2465.
[10] Thali, M. J., Yen, K., Schweitzer, W., Vock, P., Boesch, C., Ozdoba, C., ... & Dirnhofer, R. (2003). Virtopsy, a new imaging horizon in forensic pathology: Virtual autopsy by postmortem multislice computed tomography (MSCT) and magnetic resonance imaging (MRI). Journal of Forensic Sciences, 48(2), 386–403.
[11] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115.
[12] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
[13] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
[14] Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., von Tengg-Kobligk, H., ... & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiology: Artificial Intelligence, 2(3), e190043.
[15] Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. BMJ, 330(7494), 765.
[16] Seymour, C. W., Kennedy, J. N., Wang, S., Chang, C.-C. H., Elliott, C. F., Xu, Z., ... & Clermont, G. (2019). Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA, 321(20), 2003–2017.
[17] Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., ... & Zoernig, I. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Medicine, 25(7), 1054–1056.

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 6, Issue 3, July 2026 Impact Factor: 8.2

[18] International Classification of Diseases, 11th Revision (ICD-11). (2022). World Health Organization. https://icd.who.int/en
[19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.
[20] Thali, M. J., Dirnhofer, R., & Vock, P. (Eds.). (2009). The virtopsy approach: 3D optical and radiological scanning and reconstruction in forensic medicine. CRC Press.
[21] Ebert, L. C., Ptacek, W., Fürst, M., Ross, S., Flach, P. M., Thali, M. J., & Bolliger, S. A. (2012). Forensic radiology—a new subspecialty for radiologists. Diagnostic and Interventional Radiology, 18(3), 329–333.
[22] Poulsen, K., Simonsen, J., Druid, H., & Skullerud, K. (2003). Death and complications of cardiopulmonary resuscitation. International Journal of Legal Medicine, 117(4), 195–199.
[23] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
[24] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.
[25] Bajanowski, T., Brinkmann, B., & Stefanidis, E. (2006). Infant death and malpractice in forensic medicine. Forensic Science International, 163(1–2), 15–23.
[26] Ferrante, L., Cameriere, R., & Ciafone, P. (2022). Deep learning for toxicological cause-of-death classification using postmortem blood data. Forensic Science International, 337, 111361.
[27] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis. PLOS Medicine, 15(11), e1002686.
[28] Schulz, R., Kreutz, K., Bürrig, K. F., Vieth, V., & Püschel, K. (2020). Multimodal data fusion improves post-mortem interval estimation accuracy. International Journal of Legal Medicine, 134(5), 1671–1679.
[29] Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21–45.
[30] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
[31] Berner, E. S. (2007). Clinical decision support systems: Theory and practice (2nd ed.). Springer.
[32] Bright, T. J., Wong, A., Dhurjati, R., Bristow, E., Bastian, L., Coeytaux, R. R., ... & Lobach, D. (2012). Effect of clinical decision-support systems: A systematic review. Annals of Internal Medicine, 157(1), 29–43.
[33] Bolliger, S. A., Fiala, E., & Battegay, E. J. (2015). Expert-system-based forensic toxicology interpretation: A retrospective validation study. Journal of Forensic and Legal Medicine, 31, 34–39.
[34] Oehmichen, M., Auer, R. N., & König, H. G. (2006). Forensic neuropathology and associated neurology. Springer.
[35] Ahmad, M. A., Eckert, C., & Teredesai, A. (2018). Interpretable machine learning in healthcare. Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 559–560.
[36] van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. Journal of Statistical Software, 45(3), 1–67.
[37] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357