Stacked Generalization of ANN, Random Forest, and Logistic Regression for Heart Stroke Prediction

T.Swathi Priyadarshini
Department of Computer Science and Engineering,
NeilGogte Institute ofTechnology,
Hyderabad, Telangana, India.
tiggulaswathi@gmail.com

Rayabarapu Rupak Babu
Department of Computer Science and Engineering,
Keshav MemorialEngineeringCollege,
Hyderabad, Telangana, India.
rupakr31@gmail.com

Sri Chakrita Gurijala
Department of Computer Science and Engineering (AI&ML),
Keshav Memorial Engineering College,
Hyderabad, Telangana, India.
srichakritha@outlook.com

Abstract— With the help of AI, the field of modern medicine has achieved many milestones in recent years. This includes predicting heart diseases, different types of cancer during the early stages itself and even neurodivergence in various age groups. However, the pursuit of better health outcomes is a dynamic and unceasing process. Especially in early prediction of heart stroke, there is a lot of ongoing research and development. While the progress and upgrades are positive, there are still certain areas that lack more effective solutions. Our research focuses on finding better, more layered approaches of predicting heart stroke that surpass the existing ones. A lot of the existing AI-based approaches of heart stroke prediction depend on only single models or algorithms. This can limit the prediction to that model's ability and technique of capturing patterns. Since each machine learning model and algorithm, with its unique structure and working, is good at a specific type of pattern recognition, blending the abilities of diverse models can help in better and more versatile pattern recognition and prediction. In our work, we have used the stacked generalization technique. For the base models, we trained an ANN and a random forest and used logistic regression as the meta-learner. The base models were trained and tested on a publicly available dataset, which was preprocessed and class-balanced using SMOTE. The results were promising- with an increased accuracy and AUC as compared to the results of prior work, involving a custom clustering algorithm using Naïve Bayes classifier(MMAM-NB). The current work focuses on bringing an approach or solution that is more reliable in medical scenarios and has as minimal false predictions as possible.

Keywords—Heart Stroke Prediction, Ensemble Stacking, Neural Networks, Random Forests, Logistic Regression, Medical Diagnostics, Classification, Machine Learning, Model Evaluation, ROC Analysis, Data Balancing, SMOTE, Hyperparameter Tuning, Feature Optimization, Dimensionality Reduction, Visualization Techniques.

I. INTRODUCTION

Heart stroke has always been one of the most severe and life-threatening health challenges, often causing serious disability and mortality if not treated properly or detected in time. Early detection is the key to improving patient outcomes but most machine learning approaches, like Naïve Bayes, Decision Trees, and isolated neural Networks perform independently. Although these methods have shown a decent prediction accuracy, their ability to adapt to diverse patient demographics or manage imbalanced datasets is still limited, highlighting the need for better solutions.

To address this shortfall, we decided to use an ensemble stacking approach that combines the predictive capabilities of Artificial Neural Networks and an ensemble of decision trees, overseen by a probabilistic classifier as the meta-learner. This methodology also involves data normalization and Synthetic Minority Oversampling Technique (SMOTE) to remove class imbalance and reduce feature disparities. By integrating diverse modeling strengths, our method stands out from the traditional singular model frameworks, tapping into the potential of heterogeneous algorithms.

Although ensemble techniques have gained their presence in other medical domains, their exploration in heart stroke prediction has been notably sparse, positioning this study as a valuable contribution to the field. This paper outlines the complete structure of our ensemble system, assesses its effectiveness using a real-world stroke dataset, and benchmarks it against existing methods from our research group. The results emphasize the promise of stacked generalization in advancing stroke prediction, presenting a reliable and adaptable technique for AI-supported clinical decision-making.

II. RELATED WORK

The domain of predictive modeling for cardiovascular health, especially heart stroke detection, has seen a rapid increase in machine learning applications, driven by the urgent need to reduce a major cause of global mortality. Early studies focused on using individual classifiers to identify risk factors, providing preliminary insights into stroke management. Research using basic predictive algorithms laid the groundwork by examining features such as age, hypertension, and glucose levels, creating a foundation for more advanced methods. These efforts highlight the significance of preprocessing and feature selection, though they often fall short in addressing the complexities of rare stroke events—a limitation that our study aims to overcome through a clinically oriented approach.

Recent developments have increasingly adopted ensemble learning to improve prediction reliability, addressing the shortcomings of single models in capturing complex and diverse data patterns. T. S. Priyadarshini et al. [1] presented a framework that combines the Maximum and Minimum Attribute Method (MMAM) with K-means clustering, paired with Naïve Bayes (MMAM-NB) and Decision Tree (MMAM-DT) classifiers. This approach

demonstrated its effectiveness by achieving an AUC of 0.96 and 0.90 accuracy, illustrating the potential of clustering-enhanced methods for predictive modeling. However, its reliance on individual classifiers and sensitivity to the initial selection of cluster centroids may limit flexibility when applied to varied datasets.

H. Puri et al. [2] proposed a predictive model using the Support Vector Machine (SVM) algorithm, incorporating features such as age, hypertension, prior heart disease, average glucose levels, BMI, and smoking status. Their investigation of linear, quadratic, and cubic decision boundaries led to improved accuracy in both training and testing across gender-specific datasets, representing an early success in classifier-based stroke prediction. However, the reliance on SVM variants limits the model's ability to handle imbalanced data effectively, potentially affecting the detection of rare stroke cases that demand specialized techniques.

N. Sharma et al. [3] employed a deep learning approach to evaluate heart stroke risk, utilizing models such as Artificial Neural Networks (ANNs) on a well-known classification dataset with preprocessed features. Their evaluation, using metrics like recall, F1-score, precision, and accuracy, found ANN to be the best-performing model, achieving 87.95% accuracy and a 91.47% F1-score. While these results demonstrate the potential of neural networks, the absence of comprehensive strategies to address class imbalance limits generalization across diverse patient populations, indicating a need for more robust preprocessing methods.

G. S. Sailasya et al. [6] explored stroke prediction by training multiple machine learning classifiers—Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Support Vector Machine, and Naïve Bayes—using physiological features. Their results identified Naïve Bayes as the top performer, achieving roughly 82% accuracy and providing a comparative baseline. However, the study lacked ensemble methods, which could improve predictive stability, and did not adequately address the common challenge of imbalanced medical datasets.

M. Wang et al. [7] proposed an undersampling-clustering-oversampling (UCO) algorithm to address imbalanced stroke-patient data in the Intensive Care Unit, targeting heart attack prediction. Using an undersampling size of 120, they reported 70.29% accuracy and 70.05% precision with a Random Forest classifier. While this method effectively mitigates data imbalance, its moderate performance suggests limitations in capturing the full complexity of stroke-related patterns, highlighting the need for more sophisticated modeling approaches.

M. GholamAzad et al. [17] used Logistic Regression to develop a stroke risk prediction model, analyzing clinical data from 5,411 patients to identify important pathogenic factors. While this approach offered a structured framework for risk assessment, its exclusive reliance on logistic regression restricts its ability to capture non-linear relationships, potentially overlooking subtle patterns that are crucial for accurate prediction.

In recent years, several ensemble approaches have been applied to heart stroke prediction, with studies demonstrating positive results through the integration of multiple predictive models. These efforts highlight the benefits of combining diverse algorithms to improve reliability, often using mixtures like decision trees or support vector machines paired with meta-learners such as gradient boosting. However, these differ from the stacking strategy employed in our study. Our framework combines Artificial Neural Networks (ANNs) and Random Forests (RF) as base models, complemented by a logistic regression meta-classifier—a setup that appears unique within the current literature. This architectural combination is advantageous: ANNs excel at capturing non-linear patterns through their layered structures, while RFs offer a systematic approach by modeling feature interactions. The logistic regression meta-learner then integrates these outputs into a unified probabilistic prediction, supporting a more cohesive learning process. Such design variations can strongly influence the ensemble's ability to handle complex, imbalanced datasets, improve feature interpretability, and increase adaptability, paving the way for enhanced accuracy and practical utility in stroke prediction.

Collectively, these previous studies have advanced the field by examining both individual and ensemble classifiers; however, challenges remain in handling imbalanced datasets, detecting complex patterns, and emphasizing rare case identification, underscoring the need for novel approaches to extend this groundwork.

III. METHODOLOGY

A. *Data Preprocessing*

The dataset for this study consists of 1,190 patient records with 12 features, obtained from a publicly accessible medical repository. Initial analysis revealed a notable class imbalance, with stroke cases underrepresented compared to non-stroke instances. To address this, the Synthetic Minority Oversampling Technique (SMOTE) from the imblearn library was employed to generate synthetic samples and balance the classes. Furthermore, feature values were standardized using the sklearn library to ensure uniform input scales, enhancing the training stability of models sensitive to feature variance.

Fig. 1. Target Class distribution before applying SMOTE.

Fig. 2. Target class distribution after applying SMOTE.

After preprocessing, the dataset was split into training and testing sets using a 75:25 ratio..

B. Base Models

1. *Artificial Neural Network (ANN)*

The ANN, built using Keras with TensorFlow, employs a deep architecture consisting of layers with 512, 256, 128, and 64 neurons, ending with a sigmoid output for binary classification. Each layer integrates Batch Normalization and Dropout (0.3) to reduce overfitting, while ReLU activation and Gaussian noise (0.01) improve model robustness. A custom loss function combines binary cross-entropy with a penalty term (weighted 2.0) to emphasize true positives, trained using the Adam optimizer over 500 epochs with early stopping (patience=20) and a learning rate scheduler (ReduceLROnPlateau, factor=0.5). Class balance is maintained using sample weights from sklearn.utils. This configuration is particularly effective at capturing complex, non-linear patterns in risk factors, a critical advantage for handling intricate medical datasets.

2. *Random Forest (RF)*

The secondary base model employs a Random Forest classifier, implemented using the sklearn library, configured with 100 decision trees and a fixed random state to ensure reproducibility. This model excels at identifying structured, rule-based patterns, providing interpretable insights and robustness against noisy data. Unlike the Artificial Neural Network (ANN), which captures non-linear relationships, the RF relies on a tree-based framework with feature splits and hierarchical decisions, making it a complementary component to the ANN in the ensemble.

C. Stacking Ensemble

Stacking, also known as Stacked Generalization, is an ensemble technique that combines the predictions of multiple base models to form a more robust meta-model, addressing the limitations of individual models that may overfit or miss certain data patterns. While conventional single-model approaches can achieve reasonable accuracy, they often struggle to generalize across diverse medical datasets, underscoring stacking's value in enhancing reliability.

In our framework, the ANN and RF serve as base models, with their probability outputs fed into a logistic regression meta-classifier. This linear model, initialized with default settings, evaluates the contributions of both base models and applies class weights via the sklearn library to maintain balanced predictions. The ANN effectively captures complex, non-linear relationships in the data, whereas the RF emphasizes rule-based patterns and feature partitions. The logistic regression meta-classifier dynamically adjusts the influence of each base model's outputs, tailoring decisions to specific contexts, such as high-risk patient cases or ambiguous feature combinations.

We started by evaluating individual model frameworks and simple weighted averaging methods. However, these approaches showed lower efficiency and adaptability compared to stacking. For instance, the ANN had difficulty generalizing predictions across all patient groups, while weighted averaging treated all model outputs equally, ignoring the context of each prediction. In contrast, stacking dynamically learns to prioritize the more accurate prediction for each instance, offering clear practical benefits in clinical applications.

During the execution phase, the training data was split 70:30 to train the meta-classifier, transforming the outputs of the base models into meta-features. This setup enhances the ensemble's ability to maintain consistent performance across diverse patient populations.

D. Evaluation Metrics

To gauge the performance of our model, we employed metrics from the sklearn library to calculate essential indicators suited for heart stroke prediction.

- **Accuracy**represents the percentage of correctly predicted instances across the entire dataset, providing an overall indicator of model performance
- **Precision**evaluates the reliability of positive predictions, ensuring consistent and trustworthy identification of stroke cases.
- **Recall**, particularly important for rare conditions like strokes, assesses the model's ability to correctly identify true positive cases, helping to reduce the risk of missed diagnoses.
- **F1 score**, calculated as the harmonic mean of precision and recall, offers a balanced assessment, especially for datasets with imbalanced class distributions. Additionally, a detailed classification report is produced to summarize these metrics across all categories.

For visual representation, used Matplotlib to generate:

- Receiver Operating Characteristic (ROC) curves, illustrating the trade-off between true positive and false positive rates,
- Learning curves, which help assess the stability of the training process.

These quantitative metrics and graphical depictions, discussed further in the Results section, reinforce the evaluation of the stacking framework's advantages over conventional single-model approaches.

Results And Analysis

E. Quantitative Performance and Comparison

1. *Comparison Table: ANN vs. RF vs. Stacked model*

Table 1 presents the evaluation of our ensemble framework, which employed a mix of statistical metrics and visualization techniques to assess performance. The Artificial Neural Network (ANN) initially showed strong predictive capability, attaining an accuracy of 90.2% and an AUC of 0.948. This reflects its effectiveness in modeling complex, non-linear relationships within the stroke dataset, although its performance showed some variation across different patient profiles.

Next, the Random Forest (RF) classifier demonstrated robust decision-making, achieving an accuracy of 94.0% and a peak AUC of 0.983. Its hierarchical, feature-based structure produced consistent results, though it showed a minor limitation in fully capturing the complex non-linear relationships inherent in the data.

Building on these results, the meta-model, which integrates ANN and RF outputs via a logistic regression meta-learner, clearly outperformed the individual models. It achieved 94.6% accuracy, 94.1% precision, 94.7% recall, and a 94.4% F1-score, with a comparable AUC of 0.982. The near-identical AUC to RF, combined with superior balanced metrics, highlights the meta-model's ability to harness both the ANN's capacity for complex pattern recognition and the RF's decision-making stability. These findings demonstrate its value as a unified, high-performing classifier, surpassing its constituent models and establishing a strong benchmark for stroke prediction.

	Accuracy	Precision	Recall	F1-Score	AUC
ANN	0.902	0.871	0.934	0.902	0.948
RF	0.94	0.94	0.934	0.937	0.983
Meta Model(LR)	0.946	0.941	0.947	0.944	0.982

Table 1. Performance comparison of ANN, RF, and the Meta Model

2. *Classification Report of the Meta Model*

Overall performance of the Logistic Regression model:

- **Accuracy:** 94.6%
- **Precision:** 94.11%
- **Recall:** 94.73%
- **F1 Score:** 94.26%

	Precision	Recall	F1-score	Support
0	0.95	0.94	0.95	163
1	0.94	0.95	0.94	152
Accuracy			0.95	315
Macro avg	0.95	0.95	0.95	315
Weighted avg	0.95	0.95	0.95	315

Table 2. Classification report of the stacked ensemble model

Table 2 emphasizes the meta-model's balanced performance across both stroke and non-stroke classes. It consistently achieves accuracy and recall values between 0.94 and 0.95 for each class, supported by macro and weighted average F1-scores of 0.95. This consistency demonstrates the model's robustness in handling imbalanced datasets and accurately detecting stroke cases without inflating false positives. Such reliability is essential for safe and effective clinical decision-making in stroke screening.

3. *Confusion Matrices-Analysis*

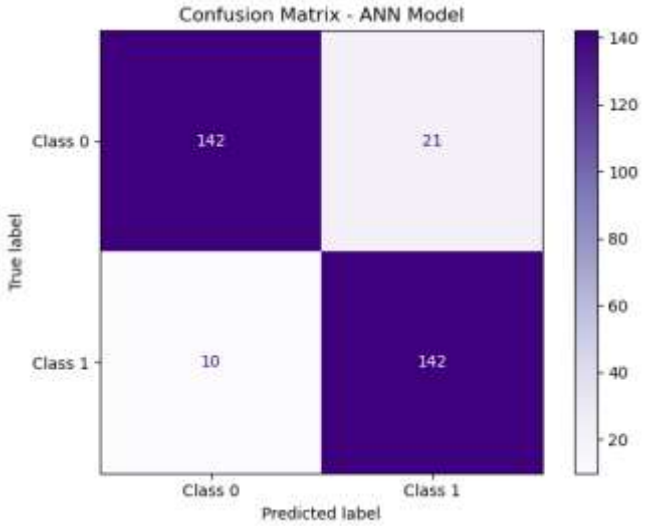

Fig. 3 Confusion Matrix of ANN

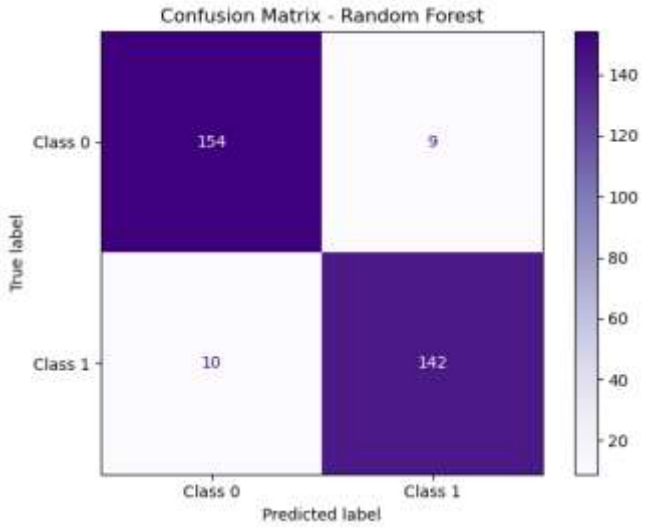

Fig.4. Confusion Matrix of Random Forest

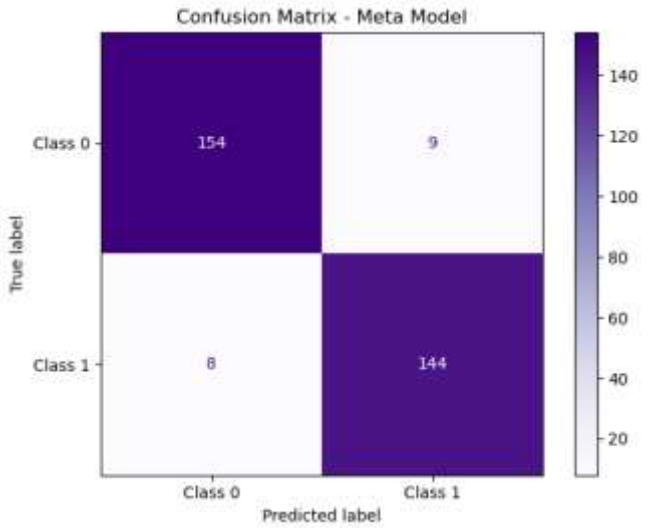

Fig. 5 Confusion Matrix of the meta model

Figs. 3–5 display the confusion matrices for the three models. The ANN model (Fig. 3) incorrectly labels 21 non-stroke cases, which is notably higher than both the RF and the meta-model. The RF model (Fig. 4) performs better but still misclassifies 10 positive stroke cases. In contrast, the stacked ensemble (Fig. 5) achieves the most balanced performance, reducing both false positives and false negatives (9 and 8, respectively), highlighting its effectiveness in high-risk scenarios such as stroke prediction.

F. Visual Evaluation of Discriminative Ability

1. ROC Curve Analysis

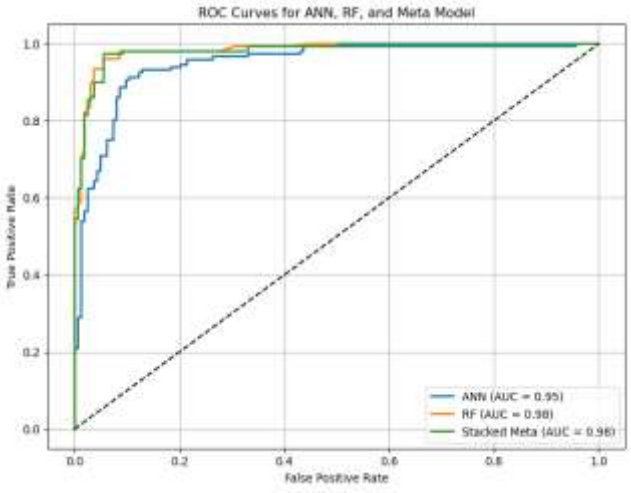

Fig. 6 ROC Curves

The Receiver Operating Characteristic (ROC) curve provides a visual method for assessing a model's ability to differentiate between positive and negative classes across a spectrum of decision thresholds. In Fig. 6, we present the ROC curves for all three models, highlighting their relative performance. Both the Random Forest (RF) and the meta-model achieve an AUC of 0.98, but the meta-model's curve shows greater stability, maintaining a more consistent balance between true positive and false positive rates across thresholds. In comparison, the Artificial Neural Network (ANN) attains a lower AUC of 0.95, reflecting its relative difficulty in handling data variability on its own.

2. Precision-Recall Curve For The Meta-Model

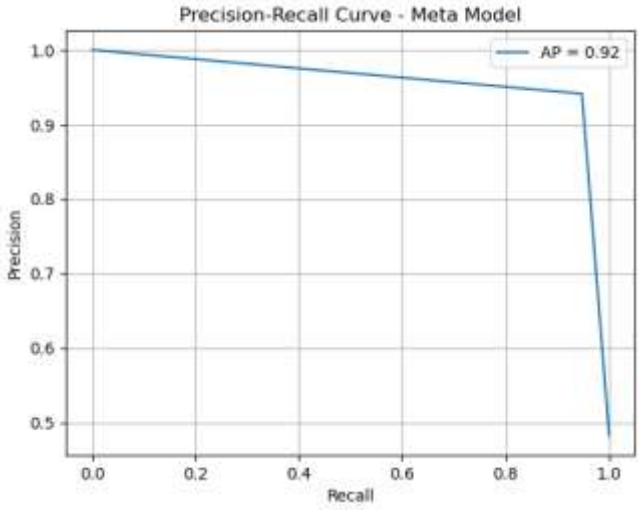

Fig. 7 Precision-Recall Curve

Given the dataset's class imbalance, the precision-recall curve offers a clearer perspective on the model's reliability in identifying stroke cases.

The Precision-Recall (PR) curve in Fig. 7 shows an Average Precision (AP) score of 0.92, highlighting the model's capability to sustain high precision across varying levels of recall. This is particularly important for imbalanced datasets such as those used in stroke prediction.

G. Model Behavior And Training Insights

1. Training Loss and Accuracy Curve- ANN

The plot in Fig. 8 illustrates the reduction in loss over epochs for both training and validation datasets. While training loss consistently declines, the validation loss levels off, suggesting limited gains in model generalization beyond a certain point. This observation supports the use of early stopping to maintain generalization and avoid additional overfitting.

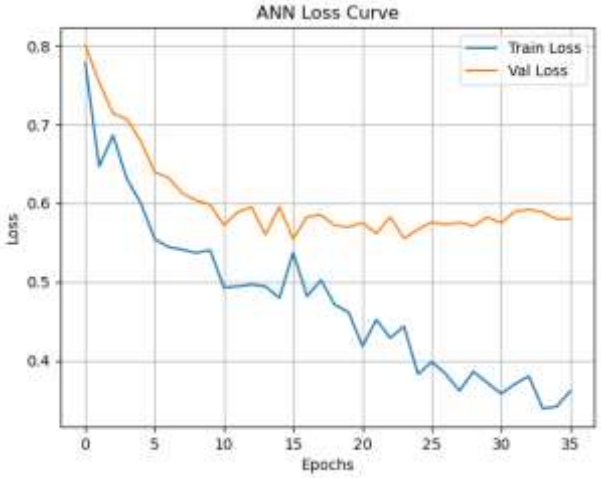

Fig. 8 Loss Curve for ANN

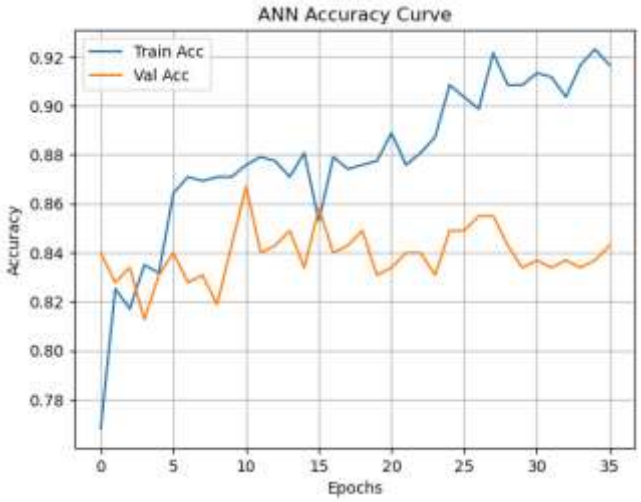

Fig. 9 Accuracy Curve of ANN

The ANN accuracy curve shows a steady increase in training accuracy over the epochs, leveling off near 92%. Although validation accuracy is lower, plateauing around 84–85%, the moderate difference indicates some overfitting, but it remains controlled due to regularization techniques such as dropout and Gaussian noise.

H. Feature Interpretability and Dimensional Insights

1. Top Feature Importances- Random Forest

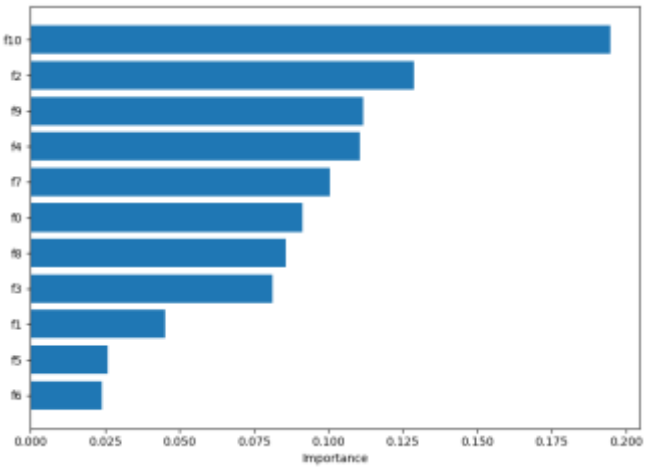

Fig. 10 Ranking of all the features by Random Forest

The bar plot in Fig. 10 ranks the most important features according to their contribution in the Random Forest model. Feature f10 emerges as the most significant, followed by f2 and f9. These results enhance model interpretability and can also assist domain experts in pinpointing key clinical markers associated with stroke risk.

2. PCA and t-SNE Projections of the Dataset

The PCA plot offers a linear projection of the data, showing overlap between the two classes but also some separability in PC1 and PC2. This supports the idea that more expressive models (like ANN and RF) are beneficial for capturing non-linear relationships in the data.

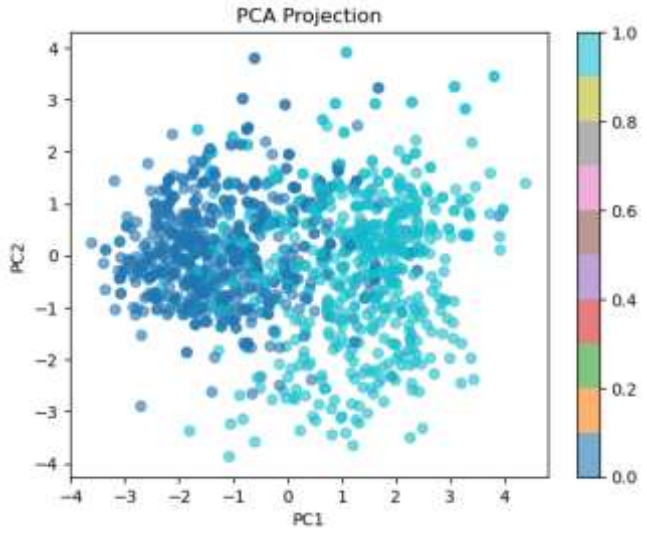

Fig. 11 PCA Projection

The t-SNE projection (Fig. 12) provides a non-linear dimensionality reduction that visually highlights clusters within the dataset. The presence of clearly separated groups suggests that the features capture meaningful patterns for classification, supporting the appropriateness of the feature space for subsequent modeling.

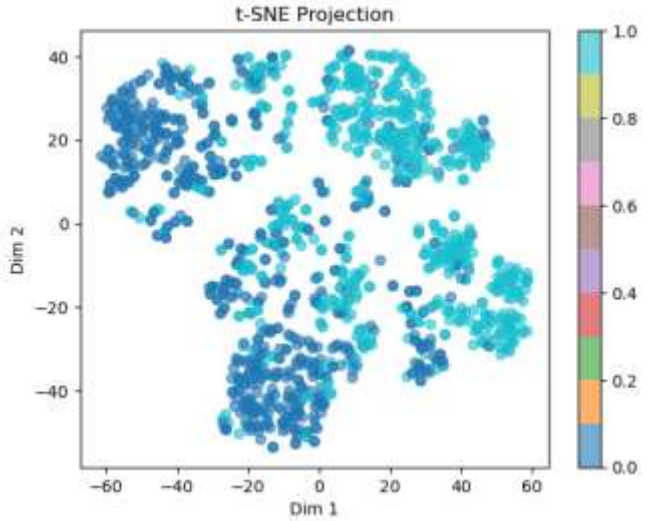

Fig. 12 t-SNE Projection

IV. DISCUSSION

The results of our stacked ensemble framework merit a thorough evaluation in the context of existing stroke prediction approaches. This section examines the comparative advantages of our method, the rationale underlying our design choices, and its prospective implications for clinical practice.

A. Comparative Performance Evaluation

The stacked ensemble, combining an Artificial Neural Network (ANN) and Random Forest (RF) as base learners with a logistic regression meta-learner, consistently outperformed individual models across all evaluated metrics. The meta-model achieved an accuracy of 94.6%, precision of 0.941, recall of 0.947, F1-score of 0.944, and an AUC of 0.982. While the RF alone reached a slightly higher

AUC of 0.983, its recall was slightly lower than that of the ensemble. Although the improvement in recall may appear modest numerically, it has considerable clinical significance. The meta-learner not only increases recall but also reduces false negatives and false positives relative to its constituent models. In stroke prediction, minimizing false negatives is crucial to avoid missed diagnoses that could delay treatment or result in severe outcomes. Similarly, reducing false positives helps prevent unnecessary anxiety, additional testing, and potential overtreatment. This balance enhances the ensemble's reliability in clinical settings, particularly where high sensitivity must be maintained without compromising specificity.

B. Rationale for Ensemble Learning

The relatively modest improvements over single models might appear subtle, but the main advantage of stacking lies in its ability to integrate multiple learning strategies. The ANN captures non-linear and complex interactions among features, whereas the Random Forest provides structured, interpretable decision boundaries based on hierarchical feature relationships. The logistic regression meta-learner then dynamically calibrates the contributions of these outputs, fine-tuning the final predictions according to the context-specific reliability of each base model.

In contrast to basic averaging approaches like soft voting, which treat all model outputs equally, stacking incorporates an additional learning layer that adaptively refines decision boundaries. This setup enhances generalization and lowers the risk of overfitting, a recurring challenge during model development. To address this, we applied regularization strategies such as dropout, batch normalization, Gaussian noise, and early stopping. These interventions were crucial in stabilizing the ANN's training process, helping the model avoid excessive fitting to the training set, as evident from the balanced confusion matrix and improved classification metrics.

C. Benchmarking Against MMAM-Based Approaches

We compared our findings to a recent method that integrates the Maximum and Minimum Attribute Method (MMAM) with Naïve Bayes (MMAM-NB) and Decision Trees (MMAM-DT) [1]. The MMAM-NB approach achieved 90% accuracy with an AUC of 0.96. By comparison, our proposed framework improved accuracy by approximately 4.6 percentage points while maintaining a comparable AUC.

The structural distinction between these approaches is considerable. MMAM-based methods rely on optimized centroid initialization within k-means clustering as a preparatory stage for classification. Although this strategy improves data partitioning, it confines the predictive capacity to relatively simple classifiers. In contrast, our ensemble framework integrates SMOTE for class rebalancing, applies feature normalization, and leverages diverse model architectures capable of identifying more complex and flexible patterns. This integrated learning process results in higher predictive accuracy and greater robustness.

D. Clinical Relevance and Generalizability

In medical diagnostics, false negatives hold particularly critical consequences. The improvements observed in recall and F1-score strengthen the case for the meta-model's suitability in real-world applications. The training curves further demonstrate stable learning behavior and the capacity to generalize across different data subsets. Importantly, the Random Forest component within the ensemble contributes interpretability, allowing for feature importance analysis, an essential factor in fostering clinical confidence and adoption.

Moreover, the logistic regression design of the meta-learner improves interpretability by creating clear decision boundaries derived from the outputs of the base models. This balance of high predictive accuracy with transparency is particularly important for integrating the framework into clinical decision-support systems.

E. Limitations and Future Directions

The study's scope is restricted by its use of a single, fixed dataset, which may narrow its applicability across different patient groups or healthcare environments. To address this, future work will expand the research to involve larger, multi-centre datasets and also include time-based or longitudinal patient records for a deeper understanding of evolving health conditions. There is also room to explore advanced meta-learners such as gradient boosting or neural attention-based models, which could further refine and optimize the ensemble's performance.

In summary, the stacking framework presented here provides both a technically reliable and clinically relevant method for stroke prediction. It goes beyond current approaches in terms of predictability, while still maintaining interpretability and adaptability, two qualities that are highly important for real-world use in healthcare systems.

CONCLUSION

In recent years, predictive modeling in cardiovascular healthcare—especially for detecting strokes—has increasingly incorporated machine learning due to its potential in tackling one of the leading global causes of death. Early studies largely relied on individual classification methods to highlight risk indicators, providing preliminary guidance for stroke prevention and management. By analyzing key attributes such as age, blood pressure, and glucose levels, these initial approaches created a baseline for more advanced methodologies. While such work has emphasized the role of preprocessing and feature selection, many of these models fall short in capturing the complexity of stroke prediction, particularly given the rarity of such events. Our study addresses this limitation by adopting a more clinically oriented and comprehensive strategy.

This research enhances stroke prediction by proposing a novel ensemble model that integrates Artificial Neural Networks and Random Forests within a logistic regression meta-learner. The motivation for this design lies in combining the ANN's ability to capture complex, non-linear relationships with the RF's structured decision-making strengths, thereby overcoming the limitations of relying on a single classifier for diverse stroke risk factors. To further strengthen the framework, extensive preprocessing steps

were applied, including SMOTE to balance class distribution, which helps mitigate data imbalance and improve overall generalization.

The evaluation results demonstrate the effectiveness of this approach: the meta-learner achieved 94.6% accuracy, a recall of 0.947, and an AUC of 0.982, outperforming MMAM-NB (90% accuracy, 0.96 AUC). The high recall reduces false negatives, a crucial factor in preventing missed diagnoses that could delay treatment and prove fatal. At the same time, fewer false positives minimize unnecessary medical interventions, offering both clinical safety and efficiency. These outcomes highlight not only strong technical performance but also practical dependability in critical diagnostic scenarios.

Although the study's use of a single dataset remains a limitation, it establishes a solid foundation for advancing AI-based healthcare solutions. Future work will focus on expanding to multi-center datasets and incorporating real-time monitoring to broaden applicability. The interpretability of the Random Forest component, paired with the adaptability of the meta-learner, ensures clinical trust and transparency. Overall, this framework not only surpasses existing approaches but also sets the stage for building scalable, precise, and reliable predictive systems in medical practice.

REFERENCES

[1] T. S. Priyadarshini, M. A. Hameed and S. A. Qadeer, "NDeep Learning Heart Stroke Prediction Model Integration of MMAM with NB(MMAM-NB) and DT(MMAM-DT)," *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, Hamburg, Germany, 2023, pp. 373-380, doi: 10.1109/ICCCMLA58983.2023.10346833.

[2] H. Puri, J. Chaudhary, K. R. Raghavendra, R. Mantri and K. Bingi, "Prediction of Heart Stroke Using Support Vector Machine Algorithm," *2021 8th International Conference on Smart Computing and Communications (ICSCC)*, Kochi, Kerala, India, 2021, pp. 21–26.

[3] N. Sharma, M. K. Mishra, J. S. Chadha and P. Lalwani, "Heart Stroke Risk Analysis: A Deep Learning Approach," *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, 2021, pp. 543-598, doi: 10.1109/CSNT51715.2021.9509665.

[4] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[5] B. T. Mashi, M. Hamada, J. J. Tanimu, P. Robert and T. J. Samson, "An Ensemble Approach for Stroke Prediction," *2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, Kuala Lumpur, Malaysia, 2024, pp. 381-388, doi: 10.1109/MCSoC64144.2024.00069.

[6] G. S. Sailasya , G. L. A. Kumari, "Analyzing the Performance of Stroke Prediction Using ML Classification Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, 2021.

[7] M. Wang, X. Yao and Y. Chen, "An Imbalanced-Data Processing Algorithm for the Prediction of Heart Attack in Stroke Patients," *IEEE Access*, vol. 9, pp. 25394–25404, 2021, doi: 10.1109/ACCESS.2021.3057693.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[9] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[10] G. Biau, "Analysis of a Random Forests Model," *Journal of Machine Learning Research*, vol. 13, pp. 1063–1095, 2012.

[11] K. Pearson, "LIII. On Lines and Planes of Closest Fit to Systems of Points in Space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[12] L. van der Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 11, pp. 2579–2605, 2008.

[13] Scikit-learn Developers, "Ensemble methods," *Scikit-learn: Machine Learning in Python*, [Online]. Available: https://scikit-learn.org/stable/modules/ensemble.html.

[14] P. Kunwar and P. Choudhary, "A Stacked Ensemble Model for Automatic Stroke Prediction Using Only Raw Electrocardiogram," *Intelligent Systems with Applications*, vol. 17, p. 200165, 2023.

[15] P. Chakraborty et al., "Predicting Stroke Occurrences: A Stacked Machine Learning Approach with Feature Selection and Data Preprocessing", *BMC Bioinformatics*, vol. 25, no. 1, p. 329, 2024.

[16] M. C. Das *et al*., "A comparative study of machine learning approaches for heart stroke prediction," *2023 International Conference on Smart Applications, Communications and Networking (SmartNets)*, Istanbul, Turkiye, 2023, pp. 1-6, doi: 10.1109/SmartNets58706.2023.10216049.

[17] M. GholamAzad, J. Pourmahmoud, A. R. Atashi, M. Farhoudi and R. D. Anvari, "Predicting of Stroke Risk Based on Clinical Symptoms Using the Logistic Regression Method," *International Journal of Industrial Mathematics*, vol. 14, no. 2, 2022.

[18] E. Dritsas and M. Trigka, *"Stroke risk prediction with machine learning techniques,"Sensors*, vol. 22, no. 13, p. 4670, 2022.

[19] C. Sharma, S. Sharma, M. Kumar and A. Sodhi, "Early Stroke Prediction Using Machine Learning," *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Chiangrai, Thailand, 2022, pp. 890-894, doi: 10.1109/DASA54658.2022.9765307.

[20] B. M. Elbagoury, L. Vladareanu, V. Vlădăreanu, A. B. Salem, A.-M. Travediu, and M. I. Roushdy, "A hybrid stacked CNN and residual feedback GMDH-LSTM deep learning model for stroke prediction applied on mobile AI smart hospital platform," *Sensors*, vol. 23, no. 7, p. 3500, 2023. [Online]. Available: https://doi.org/10.3390/s23073500

[21] Mohan, R.N.V.J., Raju, B.H.V.S.R.K., Sekhar, V.C., & Prasad, T.V.K.P. (Eds.). (2025). Algorithms in Advanced Artificial Intelligence(1st ed.). CRC Press.

https://doi.org/10.1201/9781003641537