

CC BY-NC-ND 4.0 DEED Attribution-NonCommercial-NoDerivs 4.0 International

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Received 6 December 2023, accepted 14 December 2023, date of publication 25 December 2023,
date of current version 28 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3346764

RESEARCH ARTICLE

Enhancing Computer-Aided Cervical Cancer Detection Using a Novel Fuzzy Rank-Based Fusion

PRANAB SAHOO¹, SRIPARNA SAHA¹, (Senior Member, IEEE),
SAMRAT MONDAL¹, (Senior Member, IEEE), MANJEEVAN SEERA²,
SAKSHAM KUMAR SHARMA³, AND MANISH KUMAR⁴

¹Computer Science and Engineering Department, Indian Institute of Technology Patna, Bihta, Patna 801106, India

²Econometrics and Business Statistics, School of Business, Monash University Malaysia, Subang Jaya, Selangor Darul Ehsan 47500, Malaysia

³University of Maryland, Baltimore County, Baltimore, MD 21201, USA

⁴National Institute of Technology Mizoram, Aizawl 796012, India

Corresponding author: Pranab Sahoo (pranab_2021cs25@iitp.ac.in)

ABSTRACT Cervical cancer is a severe and pervasive disease that poses a significant health threat to women globally. The Pap smear test is an efficient and effective method for detecting cervical cancer in its early stages. However, manual screening is labor-intensive and requires expert cytologists, leading to potential delays and inconsistencies in diagnosis. Deep Learning-based Computer-Aided Diagnosis (CAD) has shown significant results and can ease the problem of manual screening. However, one single model is sometimes insufficient to capture the complex data pattern for accurate disease prediction. In this work, we develop an end-to-end architecture utilizing three pre-trained models for the initial cervical cancer prediction. To aggregate the outcomes of these models, we propose a novel fuzzy rank-based ensemble considering two non-linear functions for the final level prediction. Unlike simple fusion techniques, the proposed architecture provides the final predictions on the test samples by considering the base classifier's confidence in the predictions. To further enhance the classification capabilities of these models, we integrate advanced augmentation techniques such as CutOut, MixUp, and CutMix. The proposed model is evaluated on two benchmark datasets, SIPaKMeD and Mendeley LBC, using a 5-fold cross-validation approach. On the SIPaKMeD dataset, the proposed ensemble architecture achieves a classification accuracy of 97.18% and an F1 score of 97.16%. On the Mendeley LBC dataset, the accuracy reaches 99.22% with an F1 score of 99.19%. Experimental results demonstrate the proposed architecture's effectiveness and potential in cervical Pap smear image classification. This could aid medical professionals in making more informed treatment decisions, improving overall effectiveness in the testing process.

INDEX TERMS Cervical cancer, deep learning, fuzzy rank based ensemble, Pap smear test.

I. INTRODUCTION

Cervical cancer ranks as the fourth most prevalent disease affecting women, comprising 6 to 10% of all cancer cases globally [1]. The primary contributor to cervical cancer is the human papillomavirus (HPV), a sexually transmitted infection [2]. The timely identification of cervical cancer is pivotal for successful treatment, given its heightened

treatability in the early phases of diagnosis. However, if left undetected and untreated, it can metastasize to other body parts, posing a significant threat to life. The Pap smear, a standard screening test, collects the cells from several cervix regions and examines them under a microscope to identify precancerous cells that may later grow into a malignancy. However, evaluating precancerous cells is a time-consuming process that demands the expertise of qualified physicians. This intricate procedure involves analyzing patterns and meticulously classifying each cell on a slide,

The associate editor coordinating the review of this manuscript and approving it for publication was Roberta Palmeri¹.

often dealing with thousands of cells, making it a laborious and meticulous task. Moreover, variations in assessment among individuals or potential inaccuracies in prognosis due to human error can compromise patient well-being and, in rare cases, lead to adverse outcomes. The advancements in deep learning (DL) have led to the development of several computer-aided diagnoses (CAD) systems in the healthcare domain [3], [4], [5]. Numerous techniques have been devised over time for the automated detection of cervical cancer through the analysis of cytology images. Plissiti et al. [6] first introduced the SIPaKMeD Pap smear dataset and extracted the differentiating descriptors from the cell using texture, intensity, and shape-based features, followed by the SVM classifier, and reported an accuracy of 95.35%. In a separate study [7], the authors utilized the Random Forest algorithm and implemented a hybrid approach, integrating RF with shallow neural networks. Basak et al. [8] introduced a novel system that leverages multiple pre-trained models to extract deep features. Subsequently, Principal Component Analysis (PCA) and Grey Wolf Optimizer (GWO) techniques are applied to reduce the feature space's dimensionality effectively. The authors in [9] introduced a two-stage architecture to extract the texture information followed by a classifier. Researchers also utilized DL-based approaches like Orhan Yaman et al. [10] proposed a novel pyramid-deep architecture containing two steps and used DarkNet19 and SVM to classify the cervical images. Kundu et al. [11] employed two pre-trained CNN as feature extractors and applied a Genetic Algorithm (GA) for feature selection followed by SVM for classification. The authors in [12], introduced DeepCELL, specifically designed to classify cervical cytology images by learning feature representations through several kernels of different sizes, contributing to its effective image classification capabilities. Vision Transformer (ViT)-based approaches have also shown state-of-the-art (SOTA) results in the medical image classification problems [13], [14], [15], [16], [17]. The authors in [18] used ViT and DenseNet161 to classify cervical cytology images and attained 68% accuracy. Pacal et al. [19] addresses the issues of data quality and image variability by employing several advanced deep-learning techniques. They employed more than 40 CNN and 20 ViT-based models on the SIPaKMeD pap-smear dataset and reported the ViT models' superior performance with data augmentation and ensemble learning. This study in [20] introduces a novel CAM-VT framework for identifying cervical cancer nest images under weak supervision. CAM-VT integrates Conjugated Attention Mechanism and Visual Transformer modules, combining local and global feature extraction, and incorporates ensemble learning to enhance identification capabilities. They have employed one private dataset and reported an average accuracy of 88.92%. The study in [21] introduces a weakly-supervised approach for grading cervical dysplasia directly on whole-slide images. The method involves epithelium segmentation and

a dysplasia classifier, eliminating the need to identify epithelial areas manually. The proposed classification attains a balanced accuracy of 71.07% and a sensitivity of 72.18% in slide-level testing on 2000 WSIs of LEEP samples and surgical specimens collected from 1051 patients. Indeed, developing deep learning (DL)-based models typically demands a large dataset to achieve satisfactory performance, which can be challenging to obtain in healthcare settings due to privacy and data scarcity concerns. Transfer learning (TL) emerges as a viable solution to this problem. TL involves leveraging a pre-trained model trained on a large dataset and reusing it for the specific task at hand with limited data available. By utilizing the knowledge and feature representations from the extensive dataset, TL enables the model to adapt and perform effectively on the smaller, problem-specific dataset. However, different models might work well on certain data distributions, making some classes in the dataset have better categorization than others. Ensemble learning is a solution that combines the decision scores from different models to predict the final class by capturing the salient features of each of its constituent models. It typically aims to reduce bias and variance, where bias refers to the systematic error or deviation from the actual value, and variance refers to the variability of model predictions. Ensemble learning reduces variance by averaging or combining multiple predictions, improving reliability and generalization [22], [23], [24]. The authors in [25] proposed an ensemble technique combining pre-trained models using the Sugeno fuzzy integral. This method aims to enhance the predictive power and robustness of the overall model by aggregating the outputs of multiple pre-trained models through the fuzzy integral. Another author [26] utilized Inception V3 [27], MobileNet V2 [28], and Inception-ResNet V2 [27], along with some additional layers for image categorization. They utilized a fuzzy distance-based ensemble technique multiplying the distances associated with each class and selected the minimum value as the final output. Several other authors [24], [29], [30], [31] also used ensemble techniques in different medical image analysis problems and reported better performance. In recent studies, researchers have explored several simple fusion approaches to combine the outputs of different models and enhance overall performance. These methods include weighted averaging, stacking, majority voting, and others [32], [33], [34], [35]. Nevertheless, these approaches do come with certain drawbacks. For instance, Simple weighted averaging assumes that all base models have equal contributions, which may not always be the case, leading to suboptimal results. Additionally, giving higher weight to a model solely based on its accuracy may overlook the potential benefits of combining multiple models with complementary strengths [36], [37]. Stacking also requires training an additional meta-model to combine the predictions of the base models. This can be computationally expensive and may not always improve performance, whereas majority voting is sensitive to noise and outliers, as it relies on the majority of

base models agreeing on a prediction. In this proposed work, we employed three pre-trained CNNs and fine-tuned them to classify the input image into one of the classes. The selection of these pre-trained models is based on a comprehensive literature survey [12], [26], [38] and initial experiments. In the initial experiments, we employed several deep-learning models for the base classifier prediction and considered those models that show higher performance and consistency on multiple executions. We excluded models that exhibited poor performance in the first stage, as their inclusion could potentially diminish the overall performance of the entire architecture. We have utilized SE-inception-v4 [27], ResNet-152V2 [39], and DenseNet169 [40], followed by a novel fuzzy rank-based fusion technique to fuse the decision scores generated by the base classifiers. By considering rankings instead of direct class labels, it can mitigate the impact of outliers and noisy predictions, leading to more reliable ensemble predictions. Rank-based ensembles provide a clear and intuitive interpretation of the ensemble's predictions. The rankings represent the relative importance or consensus of each class across multiple classifiers, allowing for a better understanding of the ensemble's decision-making process. We have utilized two non-linear functions, namely the exponential function (exp) and sigmoid, to calculate the ranks from the predictions. The reason for employing two ranks is to consider how close or far the preliminary classification results from the predicted outcomes [41]. The term “fuzzy” in fuzzy rank-based ensemble refers to uncertainty or fuzziness associated with the rank assignments provided by individual classifiers. It acknowledges the fact that different classifiers may have varying degrees of accuracy or reliability in their rank assignments [42], [43]. We have further improved the base model's performance by incorporating advanced data augmentations, such as CutOut, MixUp, and CutMix, in addition to standard augmentations. We have validated the effectiveness of the proposed model on the publicly available SipakMed dataset [6]. The proposed model's robustness was further evaluated through additional tests on publicly available Mendeley LBC datasets. The experimental results revealed that the suggested approach outperformed several state-of-the-art techniques in the literature. A comparative analysis of the related works is shown in Table 1.

A. CONTRIBUTIONS

The key contributions of the paper are as follows:

- This study proposes an automatic screening technique of a 5-class Pap smear cervical cancer classification task by employing an ensemble of three pre-trained CNNs, serving as the base model.
- We have proposed a novel fuzzy rank-based ensemble approach by employing two non-linear functions of different concavities to determine the fuzzy ranks of the classes.
- We have employed advanced augmentation techniques like MixUp, CutMix, and CutOut with normal

augmentation during the training stage to improve the robustness of the architecture.

- We performed several experiments to evaluate the proposed model on two publicly available cervical datasets and obtained competitive results on important metrics.

The subsequent sections are organized as follows: Section II presents the problem formulation, where the problem statement is precisely defined. Section III elaborates on the proposed methodology, providing detailed insights into the approach adopted for cervical cancer detection. Section IV focuses on the dataset utilized in the experiments, outlining its characteristics, and presents the results obtained from the conducted experiments. Section V offers a comprehensive analysis of the obtained results, emphasizing the strengths and limitations of the proposed method. Finally, Section VI concludes the paper by summarizing the key findings and proposing potential future directions to enhance further cervical cancer detection using computer-aided methods.

II. PROBLEM FORMALIZATION

The primary objective of this study is to develop a novel architecture that addresses the challenges in computer-aided cervical cancer classification. First, the input images are classified into one of the classes using three different pre-trained CNNs, such as SE-Inception-V4, ResNet-152 V2, and DenseNet169, followed by a novel fuzzy rank-based fusion technique. In this rank-based ensemble approach, predictions from each model are assigned varying importance, with their weights determined by their respective ranks, improving the overall classification accuracy. Formally, consider the input dataset $D = \{D^i\}_{i=1}^N$, where N represents the total number of samples in the dataset. For a particular input sample D_i , the predictions obtained from the three pre-trained models are presented as $\{P^i\}_{i=1}^3 = [\{C_1(D^i)\}, \{C_2(D^i)\}, \{C_3(D^i)\}]$, where C_1 , C_2 , and C_3 are the three pre-trained classifiers, and P^i is the prediction of each classification model. The final output from the proposed architecture is mathematically formulated as $X = \text{RBF} \{P^i\}_{i=1}^3$, where RBF is the proposed rank-based fusion, and X is the predicted class.

III. METHODS AND MATERIALS

This section outlines the key components of the proposed architecture, followed by the advanced augmentation technique and the proposed novel fuzzy rank-based ensemble.

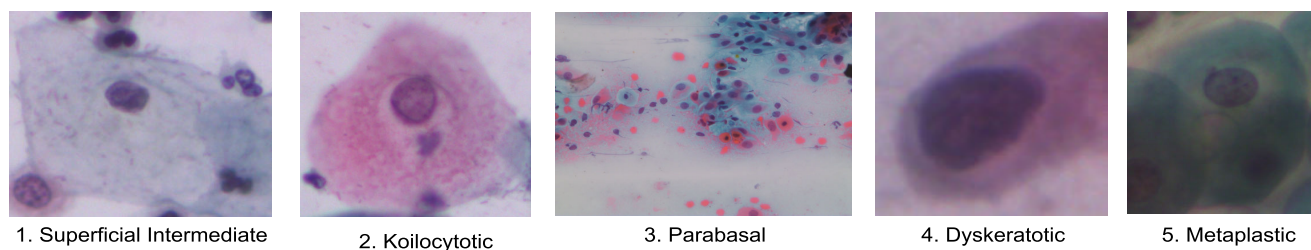
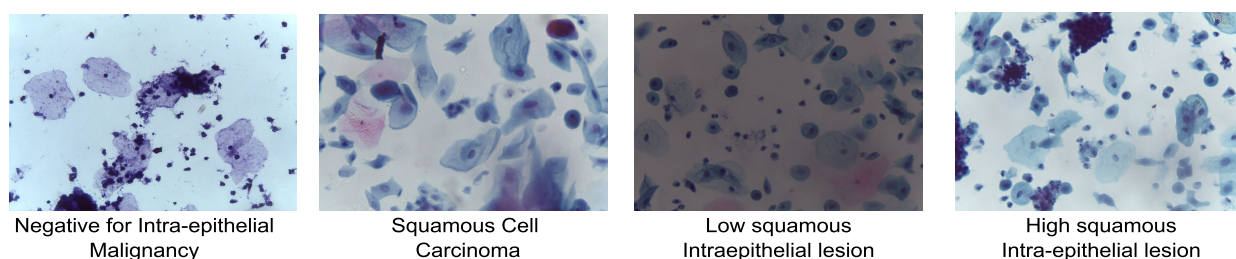
A. CLASSIFICATION MODELS

We have utilized three pre-trained models for the initial predictions: SE Inception V4 [27], ResNet-152V2 [39], and DenseNet169 [40]. The brief details about the base learners are as follows:

SE Inception-V4 is a modified variant of the Inception architecture that integrates the Squeeze-and-Excitation (SE) module. The SE model enhances the network by explicitly

TABLE 1. A summary of the related works.

Ref.	Publication Year	Dataset	Number of Images	Method	Result Accuracy (%)
Plissiti et al. [6]	2018	SIPaKMeD	4049	SVM	95.35 (5-class)
Martinez et al. [44]	2020	Private dataset	450	CNN with Cell Merger	88.8 (4-class)
Manna et al. [38]	2021	SIPaKMeD	4049	Ensemble of CNN Models	95.43 (5-class)
Chen at al. [45]	2021	Harlev	917	Knowledge distillation	73.58 (7-class)
Mehmood [7]	2021	UCI	858	Random Forest	93.6 (2-class)
Kundu et al. [25]	2021	SIPaKMeD Mendeley LBC	4049 963	Sugeno Fuzzy based Ensemble	96.33 (5-class) 99.48 (4-class)
Liu et al. [46]	2022	SIPaKMeD CRIC	8838	CNN with Visual Transformer	91.72 (11-class)
Mohammed et al. [47]	2022	SIPaKMeD Harlev	4049 917	CNN with Multiple sized kernels	95.628 (5-class) 92.717 (4-class)
Fang at al. [12]	2022	SIPaKMeD Harlev	4049 917	CNN with Multiple sized kernels	95.628 (5-class) 92.717 (4-class)
Yaman et al. [10]	2022	SIPaKMeD Mendeley LBC	4049 963	Transfer learning With SVM	96.71 (5-class) 99.47 (4-class)
Habtemariam et al. [48]	2022	JUMC	915	EfficientNetb0	96.84 (3-class)
Pramanik et al. [26]	2022	SIPaKMeD	4049	Fuzzy-distance based ensemble	96.96 (5-class)
Pacal et al. [19]	2023	SIPaKMeD	4049	ViT based Max-voting ensemble	92.95 (5-class)

**(a) SIPaKMeD Dataset****(b) Mendeley Dataset****FIGURE 1.** Example of the sample images from each datasets.

modeling the interdependencies between channels within each layer. The SE module consists of two main steps: squeezing and exciting. In the squeezing step, the spatial dimensions of the feature maps are reduced to a global spatial context by applying global average pooling. In the exciting step, the channel-wise statistics are fed into a small fully connected neural network. The SE module has a large number of filters and layers. This enables the model to learn more

complex and abstract features from the data, enhancing its representational capacity.

ResNet-152V2 is an advanced version of the original ResNet architecture specifically designed with 152 layers, making it a very deep and powerful network [39]. ResNet-152V2 utilizes bottleneck blocks to reduce computational complexity. These blocks consist of three convolutional layers, namely a 1×1 convolution, a 3×3 convolution,

and another 1×1 convolution. The first 1×1 convolution reduces the number of input channels, while the second 1×1 convolution restores the original number of channels. The 3×3 convolutions can achieve a good balance between model size and performance. ResNet-152V2 incorporates batch normalization, global average pooling, and fully connected layers at the end for classification purposes.

DenseNet169 has a 169-layered architecture, making it a deeper and more complex network than its predecessor, DenseNet121. The approach behind DenseNet is to overcome the challenges posed by the vanishing gradient problem and enhance information flow within the network. DenseNet169 promotes the efficient flow of information throughout the network and enables each layer to access and utilize information from all preceding layers. This dense connectivity facilitates gradient propagation, encourages feature reuse, and enhances the network's ability to capture complex representations from the input data. DenseNet169 incorporates bottleneck layers within each dense block. Bottleneck layers consist of 1×1 convolutional layers that reduce the number of input feature maps before the subsequent 3×3 convolution operation. Transition layers in DenseNet169 are responsible for downsampling feature maps and reducing the spatial dimensions. These consist of the 1×1 convolutional layer followed by average pooling with a reduction factor.

1) REGULARIZATION VIA ADVANCED IMAGE AUGMENTATION

Image augmentation is a popular approach for regularizing the network in supervised learning. Normal augmentation involves basic transformations, while advanced augmentation involves more sophisticated techniques that can improve the robustness and generalization of the model. Advanced augmentation involves more complex transformations, such as Cutout, CutMix, and Mixup. These techniques are designed to simulate more realistic scenarios that the model may encounter. We have used `rotation_range=40`, `shear_range=0.2`, `zoom_range=0.2`, `horizontal_flip=True`, and `vertical_flip=True` as normal augmentation. We employ these augmentations only during training to avoid overfitting problems. The details of these three augmentation strategies are described below.

CutOut [49] involves randomly removing a square-shaped pixel patch from an image during the training process. This technique improves the robustness and generalization by forcing the model to learn features from different image parts. During training, a random square patch of pixels is selected from an image and removed by setting its values to zero. During training, the cutout operation is repeated for each training data batch, with the location and size of the cutout patch randomized for each image. This augmentation process promotes the model's capacity to learn more meaningful and invariant representations, leading to better performance and improved generalization capabilities.

MixUp [50] blends two different images and their corresponding labels from the training datasets to create augmented training samples. Specifically, for each training sample, a random pair of images and their labels are selected. The images and labels are then mixed together using a mixing coefficient (λ), which is a random value between 0 and 1. This process improves the generalization and robustness of the model by encouraging it to learn more meaningful and smooth decision boundaries. The mixed image and label can be expressed using Eq. 1 and Eq. 2.

$$\text{Mixed_image} = \lambda \times \text{Image1} + (1 - \lambda) \times \text{Image2} \quad (1)$$

$$\text{Mixed_label} = \lambda \times \text{Label1} + (1 - \lambda) \times \text{Label2} \quad (2)$$

By employing MixUp during training, the model is exposed to augmented samples between the original images. This encourages the model to learn more robust and generalizable features and reduces the risk of overfitting. MixUp has been shown to be an effective regularization technique, especially when the dataset is limited, as it creates diverse and realistic synthetic samples, enriching the training data and improving the model's performance on unseen data.

CutMix [51] is another augmentation technique that involves randomly selecting two images from the training dataset and cutting and pasting rectangular patches from one image to another while updating the corresponding labels. As a result, a new training example is created that combines information from both original images within a randomized bounding box defining the patch area. The mixing coefficient is also randomized to determine the amount of information to be mixed between the two images. It is an extension of MixUp, but instead of linearly interpolating two images, CutMix performs a more aggressive augmentation by mixing and pasting a randomly cropped portion of one image onto another. Let us combine two samples, Image1 and Image2, and randomly choose a rectangular patch (bounding box) from Image1 and paste the selected patch onto Image2. Combine the pixel values of the pasted patch and the corresponding labels from both images using the mixing coefficient λ using the following Eq. 3 and Eq. 4:

$$M_i = \lambda \times \text{Patch_from_Image1} + (1 - \lambda) \times \text{Image2} \quad (3)$$

$$M_l = \lambda \times \text{Label_from_Image1} + (1 - \lambda) \times \text{Label_from_Image2} \quad (4)$$

This process results in a new training example called M_i , which is a combination of Image2 with a portion of Image1 pasted onto it. M_l is the corresponding label for the newly created training example.

B. PROPOSED ENSEMBLE APPROACH

This research introduces a novel ensemble architecture for cervical cancer detection incorporating two distinct non-linear functions. We have utilized Eq. 5 and Eq. 6 to model the base learners' predictions. The total deviation from the expected value is computed by two functions in our approach. These non-linear functions assess the overall deviation from

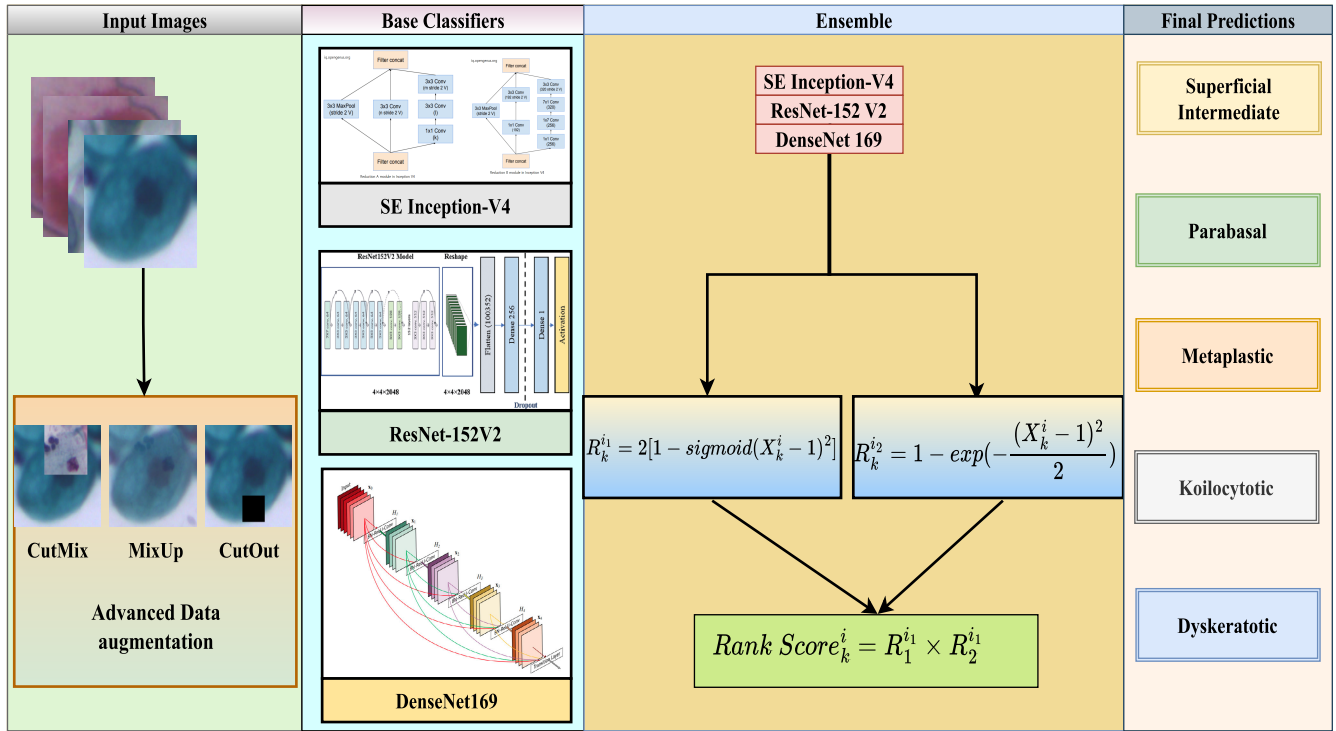


FIGURE 2. The architecture of the proposed model.

the expected value, with a lower deviation indicating greater confidence in a particular class prediction. One mapped value indicates closeness to 1, while the other indicates deviation from 1. Let $X_1^m, X_2^m, X_3^m, \dots, X_c^m$ denote the confidence scores predicted by each base model m for c number of classes, where $m = 1, 2, 3$ for three base models. After collecting the confidence scores from all the base models for a particular image, we apply Eq. 5 and Eq. 6 to map the scores. Let $R_1^{i1}, R_2^{i1}, \dots, R_c^{i1}$ and $R_1^{i2}, R_2^{i2}, \dots, R_c^{i2}$ are the fuzzy ranks generated by using Eq. 5 and Eq. 6 respectively.

$$R_k^{i1} = 2[1 - \text{sigmoid}(X_k^i - 1)^2] \quad (5)$$

$$R_k^{i2} = 1 - \exp(-\frac{(X_k^i - 1)^2}{2}) \quad (6)$$

Eq. 5 provides a reward if X approaches towards 1. Conversely, the deviation will be more in Eq. 6 if X approaches towards 0. Let $(\text{Rank_Score}_1^i, \text{Rank_Score}_2^i, \text{Rank_Score}_3^i, \dots, \text{Rank_Score}_c^i)$ be the fused rank scores, where Rank_Score_k^i is calculated using Eq. 7.

$$\text{Rank_Score}_k^i = R_1^{i1} \times R_2^{i1} \quad (7)$$

The rank score is defined as the product of reward and deviation for a specific confidence score obtained from a base learner. This score reflects the confidence level of a particular class. Finally, we calculate the fused score for k^{th} class (FS_k) using Eq. 8. The fused scores are determined by the rank scores, obtained by multiplying the fuzzy ranks generated by the non-linear functions. This fused score serves as the final

score for each class. The class with the minimum fused score is identified as the winner using Eq. 9.

$$FS_k = \sum_{i=1}^m \text{Rank_Score}_k^i, \forall k = 1, 2, \dots, c. \quad (8)$$

$$\text{class}(I) = \min[FS_k], \forall k = 1, 2, \dots, c. \quad (9)$$

The graphs of the functions are shown in Fig. 3. The Eq. 5 is concave upward within its domain $[0, 1]$. As X approaches 1, the value of sigmoid decreases making the fuzzy rank increase. Equation 6 is concave downward within its domain $[0, 1]$. When X approaches 0, the value of exp decreases making the fuzzy rank increase. The outcome of Eq. 6 being smaller than Eq. 5 means that Eq. 6 governs the nature of the product. A smaller deviation from the confidence score leads to a lower rank score. Ultimately, the rank scores are the only aspect of interest when calculating the fused scores. By observing the graph in Fig. 3, it becomes evident that the final rank decreases as the confidence (probability) score increases. This observation serves as evidence supporting the correctness of the rank generation process.

Figure 4 presents an example by taking an input image from the dataset. The probability values from three base learners for each of the five classes are collected. For an image belonging to class 2, the probability value for class 2 given by SE-Inception V4 is 0.312, resulting in corresponding ranks of 0.768 and 0.211 as calculated from Eq. 5 and Eq. 6. The rank score is then determined as 0.162 using Eq. 7. Similarly, rank scores are calculated

for each base learner for all five classes, resulting in 0.162, 0.188, and 0.113 for class 2 from SE-Inception V4, DenseNet169, and ResNet-152V2, respectively. The fused score is computed as 0.463 using Eq. 8 for class 2, and the fused scores of 0.539, 0.632, 0.564, and 0.549 are obtained for classes 1, 3, 4, and 5, respectively (as mentioned in the Fig. 4). The winner class is determined as class 2, chosen by SE-Inception V4, ResNet-152V2, while DenseNet169 selects class 1. The fusion method successfully makes a robust decision, as the overall fused score is minimal for class 2. Therefore, according to Eq. 9, the predicted class is 2, confirming the initial explanation.

IV. DATASET AND EXPERIMENTAL RESULTS

A. DATASET

We have employed the popular SIPaKMeD [6] pap smear dataset to evaluate the proposed model. The dataset includes images of five categories comprising 4,049 cropped images obtained from 966 cluster cell slides. Table 2 presents the specific details of the dataset, including its size and other relevant information. The cells in the dataset fall into two categories: normal and abnormal. Parabasal and superficial-intermediate cells are considered normal, while koilocytotic, dyskeratotic, and metaplastic cells are classified as abnormal. Sample images from the dataset are shown in Fig. 1.

TABLE 2. SIPaKMeD pap smear dataset.

Class	Category	Cell Type	Images
1	Abnormal	Dyskeratotic	813
2	Abnormal	Koilocytes	825
3	Benign	Metaplastic	793
4	Normal	Parabasal	787
5	Normal	Superficial-Intermediate	831

B. EXPERIMENTAL RESULTS

We performed three experiments to showcase the efficacy of the proposed architecture. These experiments are carried out using three different settings: normal augmentation, advanced augmentation, and a combination of both. To ensure fair evaluation, we performed 5-fold cross-validation since the SIPaKMeD dataset lacks a standard train-test split. For each fold, the dataset has been divided into five parts, with one part used for testing and the other four used for training and repeated five times. The hyperparameters for training the CNN models are carefully selected through a rigorous experimental process. All the input images are standardized to a uniform size of 224×224 pixels. The Adam optimizer is utilized due to its demonstrated effectiveness in deep learning applications [52]. An initial learning rate of 0.0001 is employed to facilitate efficient and stable model training. Cross-entropy is chosen as the loss function, as it is well-suited for classification tasks [53]. A batch size of 32 is used to balance computational efficiency with generalization

performance. Each model is trained for a maximum of 100 epochs to ensure sufficient exposure to the training data. Table 3 displays the outcomes of the individual pre-trained models, followed by the results obtained from the proposed ensemble approach using different augmentation settings. Among the pre-trained models, SE-Inception-V4 consistently outperforms other pre-trained models due to its integration of Inception modules and Squeeze-and-Excitation (SE) blocks. Inception modules extract features at diverse scales through a parallel architecture, utilizing convolutional filters with varying kernel sizes. SE blocks assess feature map significance, selectively amplifying informative features. This synergistic combination enables SE-Inception-V4 to capture intricate patterns and relationships in cervical cancer images, leading to more accurate classification. In contrast, ResNet-152 V2 consistently exhibits lower accuracy due to its high complexity and extensive data requirements. Due to its vast number of parameters, ResNet-152 V2 requires a large amount of training data to learn the intricate features crucial for accurate cervical cancer classification. The limited size of the cervical cancer dataset likely prevented the model from fully exploiting its potential, resulting in suboptimal performance. The proposed ensemble model achieved 96.93% with normal augmentation while 95.85% with advanced augmentation. In the third experiment, we combine the normal and advanced augmentation techniques and achieve an accuracy of 97.18%, outperforming the previous two approaches. The enhanced performance of the proposed ensemble models stems from their ability to dynamically adapt predictions based on the confidence scores of the individual models. This adaptive mechanism prioritizes predictions from models that exhibit greater confidence in their assessments, substantially improving overall accuracy. An important observation from all the experiments is that models trained using advanced augmentation techniques achieve lower accuracy than models trained with normal augmentation alone. However, when we combine normal and advanced augmentation techniques, there is a significant performance improvement. One key finding from the experiment is that advanced augmentation, specifically CutMix, may not be suitable for every dataset. CutMix introduces artificial patches in sample images, which can appear unrealistic and negatively affect the model's performance. Therefore, it is advisable to use a combination of both normal and advanced augmentations to achieve a performance boost while avoiding potential artifacts introduced by advanced augmentation techniques. Figure 5 shows all the pre-trained models' loss vs. epoch and accuracy vs. epoch graphs in the first fold of normal with an advanced augmentation training approach. The rank-based ensemble algorithm entails ranking individual model predictions and making a final decision based on aggregated ranks. The algorithm's complexity is expressed as $O(C * (n * \log(n)))$, where C represents the number of classes, and n is the number of classifiers. Experimental trials were conducted on a Linux-based workstation with 32GB RAM, housing an NVIDIA Quadro RTX 4000 GPU with

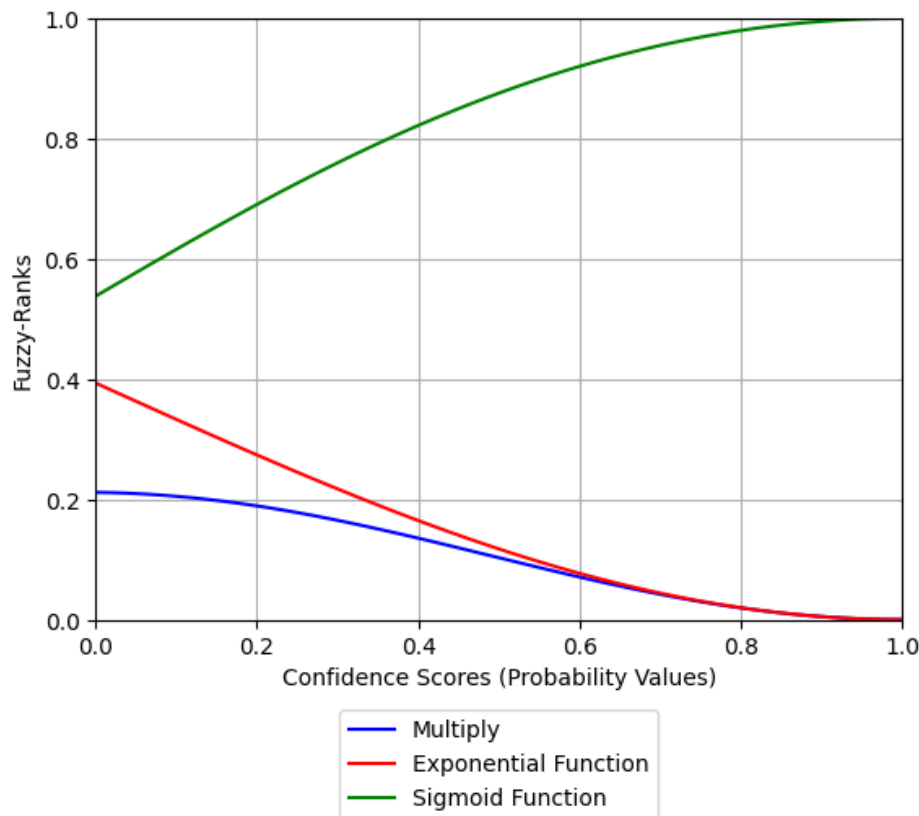


FIGURE 3. Plot of the two non-linear functions used in the proposed work and the resultant function.

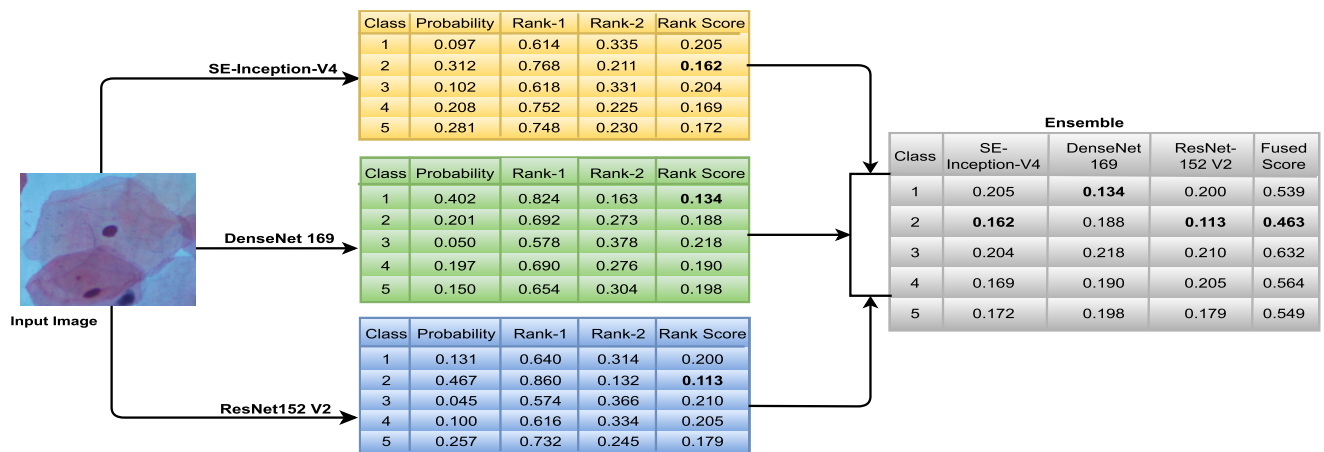


FIGURE 4. Example of the proposed ensemble approach. First, calculate the rank scores from the three base learners, then the overall fused score to get the final prediction on the classification result.

40GB memory. DenseNet169 and SE-Inception-V4 require approximately 2 hours for training, while ResNet-152V2 demands 3 hours.

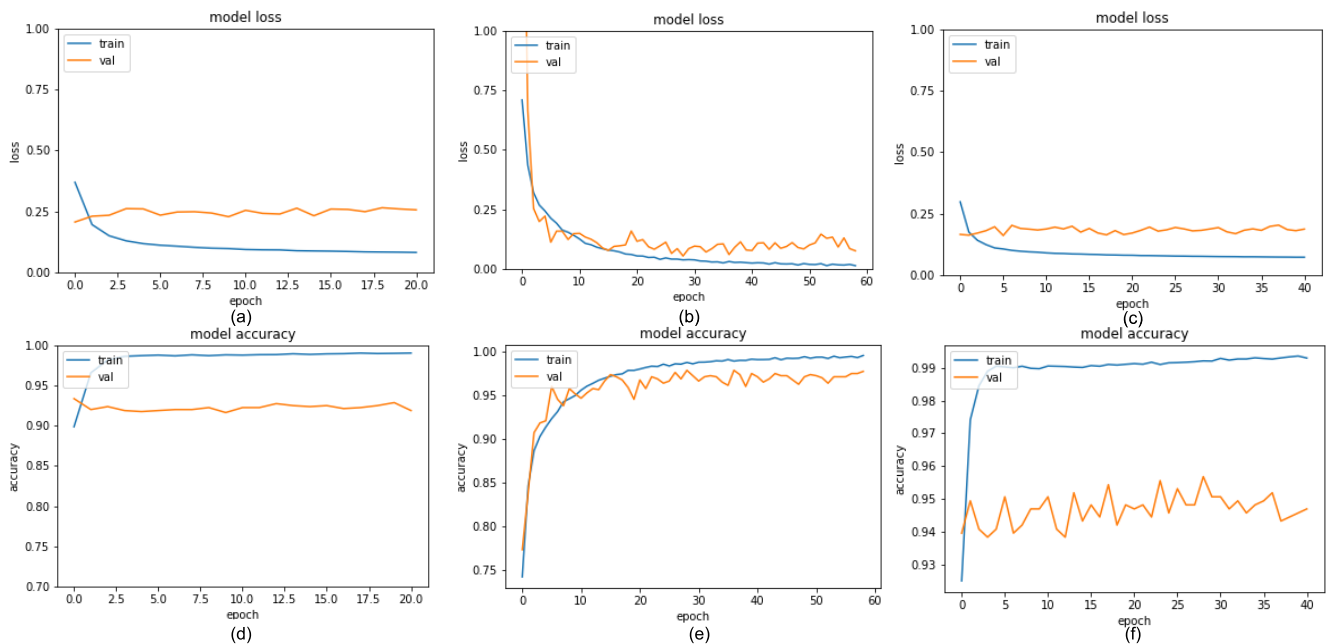
C. ERROR ANALYSIS

The presented architecture demonstrates robust and dependable performance in the task of cervical cancer classification. However, it is crucial to acknowledge potential drawbacks

when designing any new approach. We present three case studies to highlight specific instances where the proposed model encounters challenges: For example, Fig. 6(a) shows a sample from the SIPaKMeD dataset belonging to the “Dyskeratotic” class, where one individual classifier in the first stage made wrong predictions, while the proposed ensemble model has made correct predictions. Figure 6(b) shows a sample image misclassified by DenseNet169 and

TABLE 3. Experimental results of 5-fold cross-validation.

	Model	Precision	Sensitivity	F1 Score	Accuracy
Normal Augmentation	DenseNet169	94.23 ± 1.54	94.31 ± 1.55	94.16 ± 1.58	94.19 ± 1.55
	ResNet-152 V2	92.70 ± 0.75	92.79 ± 0.67	92.66 ± 0.75	92.68 ± 0.76
	SE-Inception-V4	96.13 ± 0.55	96.10 ± 0.56	96.11 ± 0.55	96.45 ± 0.55
	Proposed Ensemble Model	96.96 ± 0.52	96.94 ± 0.52	96.94 ± 0.51	96.93 ± 0.51
Advanced Augmentation	DenseNet169	91.89 ± 0.79	91.82 ± 0.76	91.73 ± 0.86	91.84 ± 0.80
	ResNet-152 V2	89.90 ± 0.69	89.86 ± 0.76	89.80 ± 0.75	89.89 ± 0.72
	SE-Inception-V4	95.28 ± 0.64	95.15 ± 0.64	95.21 ± 0.65	95.34 ± 0.64
	Proposed Ensemble Model	95.87 ± 0.39	95.83 ± 0.36	95.84 ± 0.40	95.85 ± 0.40
Normal With Advanced Augmentation	DenseNet169	94.17 ± 0.97	94.24 ± 0.96	94.18 ± 0.98	94.17 ± 0.97
	ResNet-152 V2	92.65 ± 0.58	92.67 ± 0.57	92.63 ± 0.57	92.64 ± 0.57
	SE-Inception-V4	96.91 ± 0.49	96.82 ± 0.48	96.86 ± 0.48	96.70 ± 0.49
	Proposed Ensemble Model	97.19 ± 0.32	97.14 ± 0.32	97.16 ± 0.33	97.18 ± 0.32

**FIGURE 5.** Experimental outcomes after the first fold of Normal with Advanced data augmentation approach. First row: Loss vs. Epoch curves of (a) ResNet-152V2, (b) SE-Inception-V4 (c) DenseNet169. Second Row: Accuracy vs. Epochs of (d) ResNet-152V2, (e) SE-Inception-V4 (f) DenseNet169.

ResNet-152V2, whereas SE-Inception-V4 correctly predicted it. The confidence score of SE-Inception-V4 was much higher than DenseNet169 and ResNet-152V2 in this sample instance. This resulted in the ensemble method prioritizing the SE-Inception-V4 model's decision and predicting the sample to be in the correct class. Figure 6(c) shows a case where the proposed model misclassified the image. Upon careful examination, we found that all the base models misclassified the images, which resulted in misclassification in the final model. One important point to note is that when all the base classifiers misclassify a sample in the first stage, it becomes challenging for the ensemble model to predict correctly. This idea is logical and reinforced by the observation that it is challenging for an algorithm to predict with insufficient or biased information.

D. STATISTICAL ANALYSIS

We have performed McNemar's statistical test to assess the statistical significance of the proposed ensemble framework compared to the individual base learners used in its formation. This test evaluates paired data and produces a "p-value," which indicates the likelihood that the two models are similar. A smaller p-value suggests greater dissimilarity between the models. The null hypothesis in the McNemar test is that the two classifiers are similar. To reject the null hypothesis, the p-value must be less than 0.05, indicating that the two models are statistically different. Table 4 shows that all p-values are less than 0.05, rejecting the null hypothesis. This implies that utilized base models significantly differ, highlighting the ensemble framework's effectiveness.

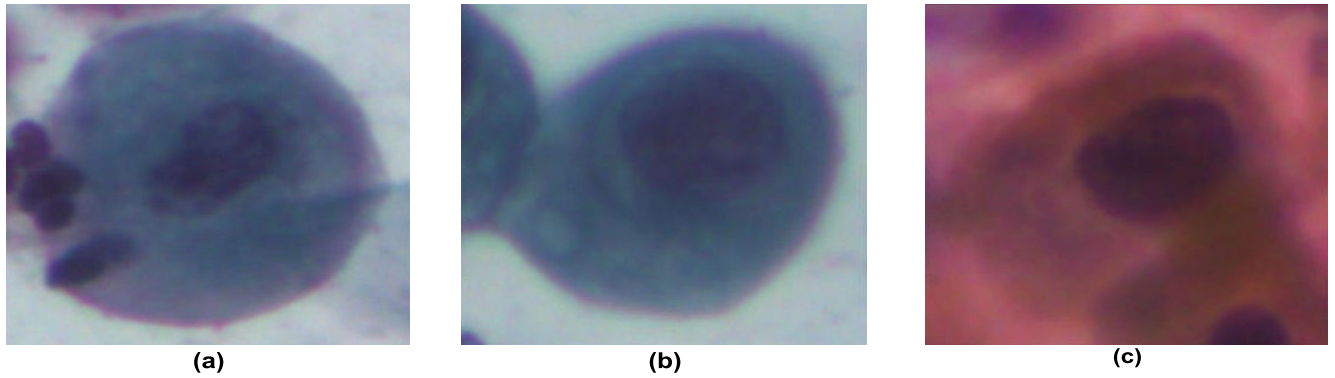


FIGURE 6. Instances of test samples are illustrated in (a) and (b) from the dataset, where individual base classifiers make incorrect predictions, but the ensemble model produces the correct prediction. Sample (c) is a case where all pre-trained models, including the proposed ensemble, misclassify the sample.

TABLE 4. Results of the McNemar's statistical test.

Model	p-value
SE-Inception-V4	3.11E-04
ResNet-152V2	3.34E-02
DenseNet169	2.06E-02

E. ABLATION STUDY

We have conducted an ablation study to demonstrate the importance of various steps in the proposed architecture. Table 3 presents the performances of individual and final models with normal, advanced, and normal with advanced augmentation. The results indicate that the proposed rank-based ensemble significantly enhances classification accuracy when employing normal with advanced augmentation.

F. ADDITIONAL EXPERIMENT ON UNSEEN DATASET

The assessment of an end-to-end AI network in medicine must be generalizable to different datasets with different patient groups. To evaluate the generalization capability of the proposed network, we conducted additional experiments on the publicly available Mendeley Liquid Based Cytology (LBC) dataset [54]. This dataset comprises 963 whole slide images of cervical cytology, distributed across four classes: Negative for Intraepithelial Malignancy (613), Low-grade Squamous Intraepithelial Lesion (LSIL) (163), High-grade Squamous Intraepithelial Lesion (HSIL) (113), and Squamous Cell Carcinoma (SCC) (74). The results in Table 5 demonstrate an impressive accuracy of 99.22% and an F1 score of 99.19% in the proposed ensemble method.

G. COMPARISON WITH STATE-OF-THE-ART APPROACHES

We have systematically compared the proposed model with several recent state-of-the-art approaches. The outcomes, depicted in Table 6, unequivocally reveal that our model outperformed the seven baseline techniques employed in the literature across all evaluated metrics. In work by Plissiti et al. [6], the authors introduced the SIPaKMeD dataset and reported an accuracy of 95.35%

using their proposed CNN architecture. Haryanto et al. [55] utilized the AlexNet architecture with a non-padding scheme and achieved an accuracy of 87.32%. Pacal et al. [19] used ViT based Max-voting ensemble and reported 92.95% accuracy in the SIPaKMeD dataset. In [56], the authors applied segmentation followed by classification using five machine learning-based classifiers and reported an accuracy of 94.09%. In another work, the authors [38] presented an ensemble technique using three pre-trained models and reported an accuracy of 95.43% with the SIPaKMeD dataset, surpassing other earlier reported approaches. Another author [10] proposed a new feature extraction technique utilizing transfer learning with an SVM classifier. Pramanik et al. [26] introduced another ensemble approach that utilized three distance measures to aggregate predictions from three pre-trained models and reported an accuracy of 97.47%. In contrast, our proposed novel rank-based ensemble approach achieved an accuracy of 97.18% with an F1 score of 97.16%, outperforming all existing SOTA baseline approaches. This improvement can be attributed to several factors. Firstly, by employing both normal and advanced augmentation techniques, our architecture enables deep learners to capture more informative features from the data, resulting in enhanced representation learning and improved classification performance. Additionally, the suggested approach incorporates a novel decision-making process. It assigns ranks to each classifier on every test sample based on the confidence of the predictions by the base models. This ranking mechanism allows for a more refined and nuanced classification process, considering the varying degrees of certainty among different classifiers. As a result, the overall classification performance is improved, leading to more accurate and reliable predictions.

V. DISCUSSION

A. GENERAL APPLICABILITY OF THE PROPOSED ARCHITECTURE

Analyzing and diagnosing cervical cancer samples is a complex process that demands careful attention to detail and a

TABLE 5. Additional experimental results on the mendeley LBC dataset.

	Model	Precision	Sensitivity	F1 Score	Accuracy
Normal With Advanced Augmentation	DenseNet169	96.56	96.92	96.72	96.32
	ResNet-152 V2	95.42	94.98	95.19	95.66
	SE-Inception-V4	98.13	97.15	97.63	98.33
	Proposed Ensemble Model	99.22	99.17	99.19	99.22

TABLE 6. Comparison with State-of-the-art methods in the literature.

Model	Precision	Sensitivity	F1 Score	Accuracy
Plissiti et al. [6]	-	-	-	95.35
Haryanto et al. [55]	-	-	-	87.32
Pacal et al. [19]	93.89	92.71	93.30	92.95
Win et al. [56]	-	-	-	94.09
Manna et al. [38]	95.34	95.38	95.36	95.43
Yaman et al. [10]	96.70	96.73	96.72	96.71
Paramanik et al. [26]	96.51 \pm 0.54	96.53 \pm 0.54	96.45 \pm 0.54	96.47 \pm 0.48
Proposed Model	97.19 \pm 0.32	97.14 \pm 0.32	97.16 \pm 0.33	97.18 \pm 0.32

comprehensive understanding of the disease's characteristics. Among the popular screening methods, the Pap smear test is crucial in detecting precancerous or cancerous cells in the cervix. This approach involves collecting cells from the cervix and analyzing them under a microscope for any signs of abnormalities that may indicate the potential development of cancer. However, there are significant challenges associated with this approach. One major obstacle is that the physical appearance of cells can change when viewed under a microscope. For example, Koilocytotic cells, which are malignant, can vary in their degeneration levels and may appear bi- or multinucleated depending on the stage of infection. Further, benign cells like metaplastic and parabasal cells, without disease, may exhibit similar physical characteristics [6]. These similarities in cell structures make visual examination complex, and there is a limited availability of testing facilities, particularly in developing nations. In this context, an automated computer-aided system can be a valuable alternative for medical professionals to enhance accuracy and efficiency in interpreting cervical images. It will assist in identifying subtle abnormalities, reducing false negatives, and improving early detection rates, contributing significantly to better patient outcomes. The proposed architecture can be extended to other cancer detection techniques like skin, breast, and lung cancer.

B. LIMITATIONS

Our study has yielded promising results in cervical cancer detection using the proposed ensemble architecture. However, challenges remain for broader applicability and improved efficiency. Firstly, the ensemble model's complexity is higher than individual architectures, which can result in longer training times. Second, deep CNN models typically require a large amount of training data to achieve optimal

performance. This is a limitation in clinical settings, where obtaining extensive datasets may not be feasible. Third, the training dataset is from one particular hospital or source, limiting the dataset's diversity and the models' generalization capability in real clinical settings.

C. FUTURE SCOPE

The proposed ensemble architecture has demonstrated promising results in cervical cancer detection, offering a potential tool for improving early detection and patient outcomes. However, further research is necessary to address the challenges of broader applicability and improved efficiency, ensuring its widespread adoption in clinical settings. Specifically, future work should focus on the following areas:

- Investigating lightweight and efficient CNN architectures within the ensemble framework could significantly reduce training time without compromising classification accuracy. These lightweight models would be particularly beneficial for resource-constrained environments, such as mobile devices or low-powered computing systems, enabling cervical cancer detection beyond traditional clinical settings.
- Exploring Meta-learning approaches could enable the model to learn from smaller datasets and perform well on diverse clinical data. Model Agnostic Meta-Learning (MAML) or Reptile can be employed to enhance the model's ability to adapt to new tasks and generalize effectively with limited training data.
- Employing federated learning techniques to train the model on distributed datasets from multiple hospitals or healthcare centers without compromising patient privacy. This would enable the model to learn from a broader range of data, improving its generalization capability in real clinical settings.

VI. CONCLUSION

Cervical cancer poses a significant health concern due to its potentially severe consequences. While medical professionals have made efforts to tackle this issue, the need for a computer-aided system for widespread screening becomes imperative to reduce the risk of human errors. This study presents a comprehensive end-to-end classification approach for detecting cervical cancer by analyzing Pap smear images. Our research introduces a novel rank-based fuzzy ensemble technique that combines multiple pre-trained models to achieve accurate predictions. At the initial prediction stage, three pre-trained models are utilized, each focusing on different aspects of the data to leverage their individual strengths. We employ two functions to aggregate the confidence scores obtained from these base models, enabling us to optimize the ensemble model's performance. By leveraging the diverse predictions and capturing complementary information provided by multiple models, the proposed rank-based ensemble improves the overall performance of the CAD system. It enables a more nuanced interpretation of the predictions, facilitating a better understanding of the decision-making process and increasing confidence in the diagnostic outcomes. The proposed framework can be seamlessly integrated as a plug-and-play model for analyzing new test images, offering clinicians faster and more accurate decision-making support. Moreover, the versatility of the proposed architecture allows for potential extension to other disease detection tasks beyond cervical cancer. As public datasets for cervical cancer are limited, future research will explore few-shot techniques in cervical cancer classification. Few-shot learning approaches can enable the model to learn from limited labeled data, thus addressing the challenge of restricted datasets and improving the model's generalization capabilities.

REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, A Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.
- [2] D. Stelzle, L. F. Tanaka, K. K. Lee, A. Ibrahim Khalil, I. Baussano, A. S. V. Shah, D. A. McAllister, S. L. Gottlieb, S. J. Klug, A. S. Winkler, F. Bray, R. Baggaley, G. M. Clifford, N. Broutet, and S. Dalal, "Estimates of the global burden of cervical cancer associated with HIV," *Lancet Global Health*, vol. 9, no. 2, pp. e161–e169, Feb. 2021.
- [3] V. Acharya, V. Ravi, T. D. Pham, and C. Chakraborty, "Peripheral blood smear analysis using automated computer-aided diagnosis system to identify acute myeloid leukemia," *IEEE Trans. Eng. Manag.*, vol. 70, no. 8, pp. 2760–2773, Aug. 2023, doi: [10.1109/TEM.2021.3103549](https://doi.org/10.1109/TEM.2021.3103549).
- [4] A. Rahman, S. Ali, A. Iqbal, S. U. Khan, A. Khan, Y. Imrana, and A. Ali, "Machine learning based parkinson's disease diagnosis using hand writing related variables," *vol*, vol. 13, pp. 143–148, Aug. 2021.
- [5] M. Guduri, C. Chakraborty, and M. Margala, "Blockchain-based federated learning technique for privacy preservation and security of smart electronic health records," *IEEE Trans. Consum. Electron.*, early access, Sep. 15, 2023, doi: [10.1109/TCE.2023.3315415](https://doi.org/10.1109/TCE.2023.3315415).
- [6] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3144–3148.
- [7] M. Mehmood, M. Rizwan, M. G. MI, and S. Abbas, "Machine learning assisted cervical cancer detection," *Frontiers Public Health*, vol. 9, Dec. 2021, Art. no. 788376.
- [8] H. Basak, R. Kundu, S. Chakraborty, and N. Das, "Cervical cytology classification using PCA and GWO enhanced deep features selection," *Social Netw. Comput. Sci.*, vol. 2, no. 5, p. 369, Sep. 2021.
- [9] S. Fekri-Ershad and S. Ramakrishnan, "Cervical cancer diagnosis based on modified uniform local ternary patterns and feed forward multilayer network optimized by genetic algorithm," *Comput. Biol. Med.*, vol. 144, May 2022, Art. no. 105392.
- [10] O. Yaman and T. Tuncer, "Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images," *Biomed. Signal Process. Control*, vol. 73, Mar. 2022, Art. no. 103428.
- [11] R. Kundu and S. Chattopadhyay, "Deep features selection through genetic algorithm for cervical pre-cancerous cell classification," *Multimedia Tools Appl.*, vol. 82, no. 9, pp. 13431–13452, Apr. 2023.
- [12] M. Fang, X. Lei, B. Liao, and F.-X. Wu, "A deep neural network for cervical cell classification based on cytology images," *IEEE Access*, vol. 10, pp. 130968–130980, 2022.
- [13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [14] P. Sahoo, S. K. Sharma, S. Saha, and S. Mondal, "A federated multi-stage light-weight vision transformer for respiratory disease detection," in *Proc. Int. Conf. Neural Inf. Process.*, Cham, Switzerland: Springer, 2023, pp. 300–311.
- [15] P. Sahoo, S. Saha, S. Mondal, and S. Gowda, "Vision transformer based COVID-19 detection using chest CT-scan images," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Sep. 2022, pp. 01–04.
- [16] R. Maurya, N. N. Pandey, and M. K. Dutta, "VisionCervix: Papanicolaou cervical smears classification using novel CNN-vision ensemble approach," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104156.
- [17] P. Sahoo, S. Saha, S. Mondal, S. Chowdhury, and S. Gowda, "Vision transformer-based federated learning for COVID-19 detection using chest X-Ray," in *Proc. Int. Conf. Neural Inf. Process.*, Cham, Switzerland: Springer, 2022, pp. 77–88.
- [18] P. Li, X. Wang, P. Liu, T. Xu, P. Sun, B. Dong, and H. Xue, "Cervical lesion classification method based on cross-validation decision fusion method of vision transformer and DenseNet," *J. Healthcare Eng.*, vol. 2022, pp. 1–10, May 2022.
- [19] I. Pacal and S. Kılıcarslan, "Deep learning-based approaches for robust classification of cervical cancer," *Neural Comput. Appl.*, vol. 35, no. 25, pp. 18813–18828, Sep. 2023.
- [20] Z. Fan, X. Wu, C. Li, H. Chen, W. Liu, Y. Zheng, J. Chen, X. Li, H. Sun, T. Jiang, M. Grzegorzczek, and C. Li, "CAM-VT: A weakly supervised cervical cancer nest image identification approach using conjugated attention mechanism and visual transformer," *Comput. Biol. Med.*, vol. 162, Aug. 2023, Art. no. 107070.
- [21] S. P. Oliveira, D. Montezuma, A. Moreira, D. Oliveira, P. C. Neto, A. Monteiro, J. Monteiro, L. Ribeiro, S. Gonçalves, I. M. Pinto, and J. S. Cardoso, "A CAD system for automatic dysplasia grading on H&E cervical whole-slide images," *Sci. Rep.*, vol. 13, no. 1, Mar. 2023, Art. no. 3970.
- [22] T. G. Dietterich, "Ensemble methods in machine learning," in *Proc. Int. Workshop Multiple Classifier Syst.*, Cham, Switzerland: Springer, 2000, pp. 1–15.
- [23] L. I. Kuncheva, *Combining Pattern Classifiers: Methods Algorithms*. Hoboken, NJ, USA: Wiley, 2014.
- [24] P. Sahoo, S. Saha, S. Mondal, and N. Sharma, "COVID-19 detection from lung ultrasound images using a fuzzy ensemble-based transfer learning technique," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 5170–5176.
- [25] R. Kundu, H. Basak, A. Koilada, S. Chattopadhyay, S. Chakraborty, and N. Das, "Ensemble of CNN classifiers using sugeno fuzzy integral technique for cervical cytology image classification," 2021, *arXiv:2108.09460*.
- [26] R. Pramanik, M. Biswas, S. Sen, L. A. D. Souza Jr., J. P. Papa, and R. Sarkar, "A fuzzy distance-based ensemble of deep models for cervical cancer detection," *Comput. Methods Programs Biomed.*, vol. 219, Jun. 2022, Art. no. 106776.

- [27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, Feb. 2017, pp. 1–14.
- [28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [29] P. Sahoo, S. Saha, S. Mondal, S. Chowdhury, and S. Gowda, "Computer-aided COVID-19 screening from chest CT-scan using a fuzzy ensemble-based technique," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2022, pp. 1–8.
- [30] M. Tanveer, M. A. Ganaie, I. Beheshti, T. Goel, N. Ahmad, K.-T. Lai, K. Huang, Y.-D. Zhang, J. Del Ser, and C.-T. Lin, "Deep learning for brain age estimation: A systematic review," *Inf. Fusion*, vol. 96, pp. 130–143, Aug. 2023.
- [31] P. Sahoo, S. Saha, S. K. Sharma, S. Mondal, and S. Gowda, "A multi-stage framework for COVID-19 detection and severity assessment from chest radiography images using advanced fuzzy ensemble technique," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 121724.
- [32] P. Zhou, X. Wang, and L. Du, "Bi-level ensemble method for unsupervised feature selection," *Inf. Fusion*, vol. 100, Dec. 2023, Art. no. 101910.
- [33] A. R. P. Syed, R. Anbalagan, A. S. Setlur, C. Karunakaran, J. Shetty, J. Kumar, and V. Niranjana, "Implementation of ensemble machine learning algorithms on exome datasets for predicting early diagnosis of cancers," *BMC Bioinf.*, vol. 23, no. 1, pp. 1–24, 2022.
- [34] V. C. Osamor and A. F. Okezie, "Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis," *Sci. Rep.*, vol. 11, no. 1, Jul. 2021, Art. no. 14806.
- [35] K. M. Sunnetci and A. Alkan, "Biphasic majority voting-based comparative COVID-19 diagnosis using chest X-Ray images," *Expert Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119430.
- [36] M. A. Ganaie, M. Hu, A. Malik, M. Tanveer, and P. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, Sep. 2022, p. 105151.
- [37] L. Nanni, S. Ghidoni, and S. Brahnam, "Ensemble of convolutional neural networks for bioimage classification," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 19–35, Jan. 2021.
- [38] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar, "A fuzzy rank-based ensemble of CNN models for classification of cervical cytology," *Sci. Rep.*, vol. 11, no. 1, Jul. 2021, Art. no. 14538.
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [40] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [41] R. Kundu, H. Basak, P. K. Singh, A. Ahmadian, M. Ferrara, and R. Sarkar, "Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans," *Sci. Rep.*, vol. 11, no. 1, Jul. 2021, Art. no. 14133.
- [42] A. F. R. de Hierro, H. Bustince, M. del Mar Rueda, C. R. De Miguel, and C. Guerra, "On the notion of fuzzy dispersion measure and its application to triangular fuzzy numbers," *Inf. Fusion*, vol. 100, Dec. 2023, Art. no. 101905.
- [43] W. Ding, H. Wang, J. Huang, H. Ju, Y. Geng, C.-T. Lin, and W. Pedrycz, "FTransCNN: Fusing transformer and a CNN based on fuzzy logic for uncertain medical image segmentation," *Inf. Fusion*, vol. 99, Nov. 2023, Art. no. 101880.
- [44] J. Martínez-Más, A. Bueno-Crespo, R. Martínez-España, M. Remezal-Solano, A. Ortiz-González, S. Ortiz-Reina, and J.-P. Martínez-Cendán, "Classifying papanicolaou cervical smears through a cell merger approach by deep learning technique," *Expert Syst. Appl.*, vol. 160, Dec. 2020, Art. no. 113707.
- [45] W. Chen, L. Gao, X. Li, and W. Shen, "Lightweight convolutional neural network with knowledge distillation for cervical cells classification," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103177.
- [46] W. Liu, C. Li, N. Xu, T. Jiang, M. M. Rahaman, H. Sun, X. Wu, W. Hu, H. Chen, C. Sun, Y. Yao, and M. Grzegorzec, "CVM-cervix: A hybrid cervical pap-smear image classification framework using CNN, visual transformer and multilayer perceptron," *Pattern Recognit.*, vol. 130, Oct. 2022, Art. no. 108829.
- [47] B. A. Mohammed, E. M. Senan, Z. G. Al-Mekhlafi, M. Alazmi, A. M. Alayba, A. A. Alanazi, A. Alreshidi, and M. Alshahrani, "Hybrid techniques for diagnosis with WSIs for early detection of cervical cancer based on fusion features," *Appl. Sci.*, vol. 12, no. 17, p. 8836, Sep. 2022.
- [48] L. W. Habtemariam, E. T. Zewde, and G. L. Simegn, "Cervix type and cervical cancer classification system using deep learning techniques," *Med. Devices: Evidence Res.*, vol. 15, pp. 163–176, Jun. 2022.
- [49] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," 2017, *arXiv:1708.04552*.
- [50] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," 2017, *arXiv:1710.09412*.
- [51] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.
- [52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [53] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8792–8802.
- [54] E. Hussain, "Liquid based cytology pap smear images for multi-class diagnosis of cervical cancer," *Data Brief*, vol. 30, 2020, Art. no. 105589.
- [55] T. Haryanto, I. S. Sitanggang, M. A. Agmalero, and R. Rulaningtyas, "The utilization of padding scheme on convolutional neural network for cervical cell images classification," in *Proc. Int. Conf. Comput. Eng., Netw., Intell. Multimedia (CENIM)*, Nov. 2020, pp. 34–38.
- [56] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. M. Aung, "Computer-assisted screening for cervical cancer using digital image processing of pap smear images," *Appl. Sci.*, vol. 10, no. 5, p. 1800, Mar. 2020.



PRANAB SAHOO received the B.Tech. degree in computer science and engineering from the Maulana Abul Kalam Azad University of Technology, West Bengal, India, in 2015, and the M.Tech. degree from IEST, Kolkata. He is currently pursuing the Ph.D. degree with IIT Patna, India. He was with AIG, India, as a Software Engineer, till 2017. His research interest includes medical image analysis using deep learning techniques.



SRIPARNA SAHA (Senior Member, IEEE) received the M.Tech. and Ph.D. degrees in computer science from the Indian Statistical Institute, Kolkata, India. She is currently an Associate Professor with the Department of Computer Science and Engineering, Indian Institute of Technology Patna, India. From September 2009 to June 2010, she was a Postdoctoral Research Fellow with the University of Heidelberg, Germany; and the Department of Information Engineering and Computer Science, University of Trento, Italy, from September 2010 to January 2011. She has more than 400 publications with 7900 citations. Her current research interests include text mining pattern recognition, natural language processing, multiobjective optimization, and biomedical information extraction.



SAMRAT MONDAL (Senior Member, IEEE) received the Ph.D. degree in computer science from the Indian Institute of Technology, Kharagpur. He is an Associate Professor with the Computer Science and Engineering Department, Indian Institute of Technology Patna, India. His publications currently report over 799 citations in Google Scholar (H-index is 17). His major research interests include security and privacy, database and data mining, and smart grid applications.



SAKSHAM KUMAR SHARMA received the B.Tech. degree in information technology from the Maharaja Surajmal Institute of Technology, Delhi, India, in 2023. He is currently pursuing the M.S. degree in computer science with the University of Maryland, Baltimore County, MD, USA.



MANJEEVAN SEERA received the Ph.D. degree in computational intelligence from Universiti Sains Malaysia. He is currently an Associate Professor of business analytics with the School of Business, Monash University Malaysia. He has over 17 years of experience in both academia and industry. His research specialization is on machine learning principles and applications in finance and engineering. His current research interests include the design and development of advanced machine

learning models for Fintech applications, particularly the detection and prediction of fraudulent financial transactions.



MANISH KUMAR is currently pursuing the bachelor's degree with the Civil Engineering Department, National Institute of Technology Mizoram.

...