A Hybrid CNN-BiGRU-GAN Framework for Enhanced Automated Analysis of Cervical Cancer in Medical Imaging

Donepudi Rohini[1], Dr. M Kavitha[2]
Research Scholar, Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation (KLEF), Vaddeswaram, Guntur Dist, Andhra Pradesh - 522302, India[1]
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India[2]

Abstract*—Cervical cancer screening requires reliable automated systems capable of overcoming variability in staining, morphology, and limited annotated data, which often undermine the performance of traditional machine learning and deep learning approaches. Existing techniques commonly rely on single-modality feature extraction or static fusion, resulting in weak generalization, class imbalance sensitivity, and limited interpretability in clinical environments. Addressing these gaps, the research introduces DiagnoFusionNet, a hybrid CNN-BiGRU-GAN framework that integrates spatial features from Convolutional Neural Network (CNN), contextual dependencies from Bidirectional Gated Recurrent Unit (BiGRU), and Generative Adversarial Network (GAN)-generated samples to enhance data diversity and correct minority-class deficiencies. The methodology incorporates an Adaptive Triple-Stage Feature Fusion mechanism that dynamically recalibrates modality contributions using discriminator-informed attention, ensuring discriminative and clinically aligned feature representations. Experiments on the SIPaKMeD dataset demonstrate strong performance with 97.89% accuracy, 97.69% precision, 96.95% recall, 96.89% F1-score, and a 0.99 AUC, supported by GAN evaluation metrics, including an FID of 18.3, IS of 3.91, and SSIM of 0.92. Ablation analysis confirms the dominant contribution of the adaptive fusion module, while t-SNE clustering and confusion-matrix inspection highlight effective separability and reduced misclassification. Model development and experimentation were executed using Python, TensorFlow, Keras, OpenCV, and Scikit-learn on GPU-enabled environments. The framework provides a clinically interpretable, data-efficient, and scalable solution for automated cervical cytology analysis in real-world and resource-limited settings.

Keywords*—Cervical cancer detection; DiagnoFusionNet; medical image analysis; Adaptive Triple-Stage Feature Fusion; generative adversarial networks

I. Introduction

Cervical cancer remains one of the most challenging global health burdens and a leading cause of cancer-related mortality among women, particularly in low- and middle-income regions where screening infrastructure is limited and access to trained cytopathologists is restricted [1]. Even though the traditional screening procedures, including Pap smears and the use of acetic acid to look into the female genitalia, have played a significant role in mortality reduction, their efficacy is limited due to the false negativities and variability in reproducibility [6] in the manual interpretation process. The differences in the quality of staining, morphological variation of cervical cells, and inter-observer variability also increase the problem of correct diagnosis, and thus, the success in detecting the disease at relatively early stages and early clinical intervention is limited significantly [2]. The growing access to digital cytology data and deep learning platforms has given rise to quicker progress in the creation of computational models that can detect fine morphological features that might not be detected with the naked eye [3], [4]. Emerging technologies in the field of artificial intelligence, especially medical imaging, promise to resolve all these gaps by offering objective, scalable, and reproducible diagnostic support [5]. The growing access to digital cytology data and deep learning platforms has given rise to quicker progress in the creation of computational models that can detect fine morphological features that might not be detected with the naked eye [6]. The natural variability of cytology images, the limited number of well-annotated data, and the need to generalize the outcome well in the diverse clinical settings, however, necessitate solutions that not only achieve good results but also can be interpreted and adapted [10]. It is against this background that the study aims to come up with an elaborate framework that will improve the reliability of the diagnostic, and the representation of the features will be reinforced [9], and is consistent with screeners' problems that are found in the real world in both resource-constrained and technologically advanced environments.

Recent developments in artificial intelligence have brought up three significant innovations that have made a significant contribution to cervical cytology analysis. Convolutional Neural Networks have been extensively utilized due to their ability to obtain rich spatial and morphological details in cervical cell images to be able to detect nuclei borders, cytoplasm texture, and structural defects [7], [8]. Although successful, these models only encode static spatial cues and do not take into consideration any contextual or temporal relationship between cellular patterns. This gap was bridged by recurrent architectures like Bidirectional Gated Recurrent Units that learned both temporal and contextual dependencies, which effectively modeled the relationships between the local structures and variations at the cell level. The classical CNN, BiGRU and GAN models are faced by the problem of data imbalance, variability of staining, and small samples. Although GANs introduce fictitious diversity, it fails to boost feature discrimination and control multimodal learning. Current hybrid approaches do not have

adaptive fusion, do not make full use of spatial, contextual, and generative advantages, generalization, interpretability, and real-world performance.

To mitigate these challenges, the proposed study provides an improved DiagnoFusionNet model that includes an Adaptive Triple-Stage Feature Fusion (ATFF) mechanism that dynamically combines spatial features of CNN, contextual features of BiGRU and synthetic diversity of GAN by an attention-driven feedback mechanism. The ATFF module facilitates the model to highlight diagnostically important morphological areas in the enhancement of robustness and generalization in the detection of early cervical cancer.

Although recent research has achieved success in deep learning-based screening of cervical cancer, most existing research employs single-modality learning (i.e., CNN-based spatial extraction, contextual modeling or GAN-based augmentation) on its own. The techniques usually apply fixed fusion algorithms, do not adaptively weight, and are not robust to class imbalance, staining variance, and sparse labels, and do not have high interpretability. Based on these gaps, this research will seek to create an adaptive multimodal model that would combine spatial, contextual, and generative representations to enhance diagnostic accuracy, generalization and clinical reliability.

A. Problem Statement

Cervical cancer remains a major cause of mortality among women in low- and middle-income countries (LMICs), where access to reliable screening technologies and trained cytopathologists is limited [10]. Existing screening approaches, such as Visual Inspection with Acetic Acid (VIA) and Pap smears suffer from high inter-observer variability, subjective interpretation, and poor scalability, while deep learning–based diagnostic systems built primarily on CNNs remain constrained by insufficient annotated datasets, severe class imbalance, and weak generalization across diverse cytological presentations. [11]. Prior studies in automated cervical cancer detection also reveal additional challenges, including limited capability to model long-range spatial–contextual relationships between cellular structures, poor robustness to noise and acquisition variations, low interpretability of model predictions, and an inability to handle the heterogeneity seen in atypical or borderline lesions [12]. Addressing these gaps requires an end-to-end architecture that simultaneously improves data diversity, balances class distributions, captures both spatial and contextual dependencies, and provides clinically interpretable outputs. To overcome these limitations, this research proposes DiagnoFusionNet—a fusion framework combining CNN-based spatial feature extraction, BiGRU contextual modeling, and a high-fidelity GAN for synthetic data generation—to deliver a more accurate, robust, and scalable solution for early cervical cancer diagnosis.

B. Research Questions

Based on the identified limitations in existing literature, this study is guided by the following research questions:

- How can spatial morphological features and contextual dependencies in cervical cytology images be jointly modeled to improve classification performance beyond conventional single-stream deep learning architectures?
- To what extent can GAN-based synthetic data augmentation alleviate class imbalance and improve model generalization in cervical cancer detection tasks?
- Can an adaptive, attention-guided feature fusion mechanism dynamically balance multimodal contributions to enhance discriminability and interpretability of diagnostic predictions?
- How does the proposed DiagnoFusionNet framework compare with classical and transformer-based models in terms of accuracy, robustness, and clinical applicability on a standard cervical cytology dataset?

C. Research Motivation

The problem of class imbalance, subtleties of morphology, and inconsistency of stains along with a lack of annotated data of high quality continue to make reliable automation of cervical cytology a challenging task. To increase diagnostic accuracy, models that are capable of identifying both fine-grained spatial information and contextual cellular interactions and are resistant to low data clinical settings are needed. To promote computer-aided screening and facilitate early detection of cervical lesions and their accuracy and resilience to imaging variability, improvements in generalization, interpretability, and resilience to imaging variability are needed in the real-world health care environment.

D. Research Significance

The study is particularly powerful as it develops AI-based cervical cytology analysis by an adaptive multimodal learning model that has the capacity to extract complementary spatial and contextual features and overcome the issue of class imbalance with samples generated with the help of GANs. Its interpretability, data efficiency and strength improve reliability and transparency of medical image assessment. This work enables early and accurate detection of the work in environments where there is limited expertise and data availability, and diagnostic consistency through combining dynamic attention mechanisms and validated fusion strategies. The strategy is scalable, clinically equivalent and is applicable to implementation in low-resource healthcare settings.

E. Key Contributions

1) Presents adaptive multimodal learning model built on spatial morphology, contextual dependency and generative augmentation to augment the robustness and interpretability of representation in cervical cytology analysis.

2) Learns an attention-directed fusion policy that learns and maintains dynamically the balance between the contributions of the features in order to enhance the discriminability, cluster separability and reliability in the presence of problematic staining variations and limited data.

3) Develops a robust evaluation process that includes ablation experiments, feature-space analysis, and GAN quality evaluation as well as analysis of errors that would ensure the validity of methods and clinical applicability.

4) Demonstrates the efficiency of the proposed DiagnoFusionNet by its higher accuracy, precision, recall, F1-score, and AUC values that are better than classical and transformer-based ones.

5) Offers ATFF-supported, GAN augmentation-supported, and empirically powerful diagnostic system, and is scalable to deployment in low-resource cervical screening offline settings.

F. Rest of the Study

Section II examines existing work. In Section III, the proposed method of working is detailed. Section IV provides findings and discussion. Conclusion and further research are given in Section V.

II. Related Works

Poudel et al. [13] implemented VIA to create a new AI-assisted method for cervical cancer detection in environments with limited resources. The study addresses the limitations of subjective VIA tests typically conducted by nurses in LMICs. They developed a protocol for quality image acquisition using smartphones, which does not require specialized devices. A dataset of 1,430 images acquired in VIA screenings was used; they planned a classification model using deep learning to classify (pre)cancerous lesions. Results show that an appropriate acquisition protocol by smartphones can deliver enough detail in images captured for VIA, and this deep-learning model was promising and could help in the actual screening with nurses and pave the way for high-volume collection and validation of data from LMICs. Though these initial studies investigated Pap-smear and VIA-based automation, they were constrained by being based on traditional classifiers or single-modality deep learning and therefore performed poorly in high-imbalanced datasets.

Wu et al. [14] analyzed the utilization of AI in cervical cancer selection, citing its promise to overcome the shortcomings of LMICs in meeting the World Health Organization's objective of screening 70% of women aged between 35 and 45 years by 2030. The study underlines the importance of accurate, efficient, and effective methods in cervical cancer screening, considering that many developing countries lack good healthcare infrastructure. Recently, AI and especially DL algorithms have proven major possible in enhancing cervical malignant disorder and aberrant cytology identification. The review is focused on the current status of AI image recognition technology for diagnosing cervical cancer, along with the challenges and possible avenues of its use in screening, treatment, and surveillance.

Borders and Thaxton Wiggins [15] proposed a study to explore Pap rates among US females, concentrating on rural-urban disparities before as well as during the COVID-19 pandemic from 2019 to 2022. The study included women between the ages of 21 and 65 and was based on nationally representative data from the "Health Information National Trends Survey". Rural females had significantly low Pap test rates compared with their urban counterparts. Rural women's unadjusted past-year Pap test rate in 2022 was 48.6%, whereas urban women's was 64.0%. With adjusted probabilities of getting a Pap test 30% lower in 2022 than in 2019, the analysis also showed a drop in Pap testing rates in 2022. This emphasizes the necessity of focused measures to raise screening rates, particularly in rural regions.

Sarhangi et al. [16] investigated how DL might improve the precision and effectiveness of cervical cancer screening, which mostly uses cytology and colposcopy. This study emphasizes that only during the stage of precancerous cervical cancer is it perfectly treated. The deep learning training process involves processes such as categorization, division, and recognition in cervical cancer diagnosis. The review also offers popular public cytology and colposcopy datasets, together with the most used deep learning architectures. Reviewing 24 practical papers, the study concludes that deep learning considerably improves the correctness and speed of cervical cancer analysis, opening doors to early diagnosis and saving lives. Whereas AI-based screening has been shown to have great promise, the majority of reported models are myopic, either utilizing spatial image features or handcrafted descriptors without considering the contextual temporal dependencies necessary to achieve adequate cell classification.

Singh et al. [17] utilized the GLOBOCAN 2020 database in assessing worldwide inequalities in incidence and cervical cancer mortality due to geographic and socioeconomic development factors. The research pursued the effort of monitoring cervical cancer towards the eradication of it from public health conditions, as recommended by the WHO, by lowering the incidence rates below 4 cases/100,000 women-years. The study reports that the occurrence and deaths from cervical cancer were significantly higher in countries with a low HDI compared to the high HDI countries. The study showed how countries with good screening services have different age-specific incidence patterns, while those with inadequate services (Colombia, India, Uganda) present an entirely different pattern.

Arbyn et al. [18] suggested a five-step framework to evaluate the possible application of biomarkers as a screening method for cervical cancer and to meet the long-felt need for proper guidelines on diagnostic research in screening situations. The framework includes: 1) correlation studies to determine the relationship between the level of neoplasia and biomarker expression, 2) clinical diagnostic studies employing the reference standard, 3) biobank-based studies that validate biomarkers by making use of archive cytology samples, 4) prospective cohort studies where baseline biomarker assessments and follow-up disease monitoring are involved, and 5) randomized intervention trials to assess reduced cancer incidence or severe dysplasia in the experimental group. This framework would help researchers to systematically evaluate new biomarkers and protect clinicians from misleading claims of tools that have not been tested well.

Shen et al. [19] designed a two-stage image synthesis system for fully automated thin-prep cytologic test systems to improve cervical abnormalities screening data. During the first phase, "Global Image Generation, a Normal Image Generator" is used to create "Cytopathological images" made up of several healthy cervical cells. Diffusion, a pretrained base model for image synthesis that has been tuned to optimize cytopathological content creation as well as spatial layout control, is the starting point for both models. Although GAN-based augmentation enhanced data diversity, the lack of incorporation with

sequential modeling constrained their diagnostic efficacy, leaving room for an even more comprehensive framework.

Li et al. [20] designed an AI approach known as STRIDE, in an effort to solve challenges in cervical cancer screening using Whole Slide Image cytology. The limitations of the current models are bypassed with STRIDE: namely, the small annotations available, as well as poor interpretability. STRIDE is a method that integrates labels at the patient level using a tiny number of cell-level annotations and uses color adversarial samples training to enhance robustness against real-world variations. Conversely, DiagnoFusionNet is the first system to synergistically combine spatial–temporal feature fusion with GAN-based augmentation in direct solution of dataset sparsity, imbalance, and interpretability issues in cervical cancer diagnosis.

Based on the literature reviewed, it is clear that past studies made use of either CNN-based spatial extraction, RNN-based contextual modeling, or GAN-based augmentation by itself. These approaches have problems of concatenating features statically, low interpretability, and generalizability when applied to unbalanced medical data. The proposed DiagnoFusionNet with ATFF overcomes the aforementioned challenges by introducing an adaptive attention-based fusion, which dynamically balances the spatial and contextual learning under the supervision of GAN, resulting in more profoundly steady, interpretable, and clinically viable cervical cancer detection.

III. Proposed DiagnoFusionNet Framework for Detecting Cervical Cancer Detection

The stated methodology presents DiagnoFusionNet, an end-to-end deep diagnostic fusion network, which is aimed at enhancing the early detection of cervical cancer by overcoming the key pitfalls of cervical cytology imaging, such as the lack of annotated data, imbalance in classes, and large cellular variations. The framework combines three complementary modules, including CNN, BiGRU, and GAN in one learning pipeline. Cervical images are first processed by resizing, normalization and augmentation using GAN. The CNN module identifies spatial morphological properties that include nuclear irregularities, chromatin texture and cytoplasmic boundaries, which are essential in the differentiation of normal and abnormal cells. The features of these spaces are then sent to a BiGRU model, which stands out by capturing contextual and sequential dependencies, therefore, allowing the model to learn some subtle cellular patterns and inter-feature relationships that cannot be easily captured by CNNs alone. In order to address the lack and unbalanced nature of data, a GAN module can be used to create high-fidelity synthetic images belonging to the minority classes, increasing the amount of data and enhancing generalization. A new ATFF mechanism dynamically integrates CNN, BiGRU and GAN-based representations with attention feedback of the discriminator, and results in a more stable and understandable diagnostic embedding. Through collaborative education of spatial, temporal, and distribution-balanced characteristics, DiagnoFusionNet improves the predictive performance, strength, and clinical viability, especially in the low-resource screening conditions.

Fig. 1 illustrates that the workflow of the DiagnoFusionNet starts with image preprocessing, which comprises of resizing, normalization, and GAN-based augmentation, then CNN spatial feature extraction, BiGRU contextual learning, and adaptive fusion. The output layer, which is the diagnostic of the fused representation, ultimately classifies the fused representation into cervical cancer categories.

A. Data Collection

The SIPaKMeD dataset [21], provided by Kaggle, one of the biggest publicly available datasets of cervical cell images, is used in the proposed study. It has 4,049 cervical cell images classified into five groups, that is, superficial-intermediate, parabasal, metaplastic, dyskeratotic and koilocytotic. The resolution of each image is 2048 x 1536 pixels, and requires preprocessing, such as resizing, normalization and augmentation, to standardize input dimension to model training. This data allows the suggested deep learning system to identify precancerous and cancerous alterations with high precision to facilitate the early diagnosis and automatic analysis of cervical cytology.

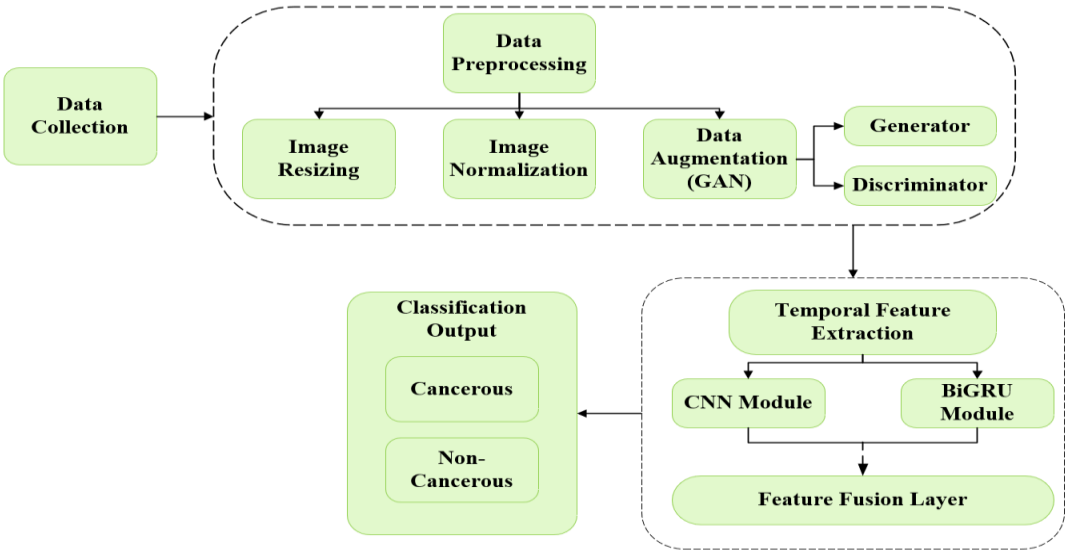

Fig. 1. DiagnoFusionNet's workflow for classifying cervical cancer.

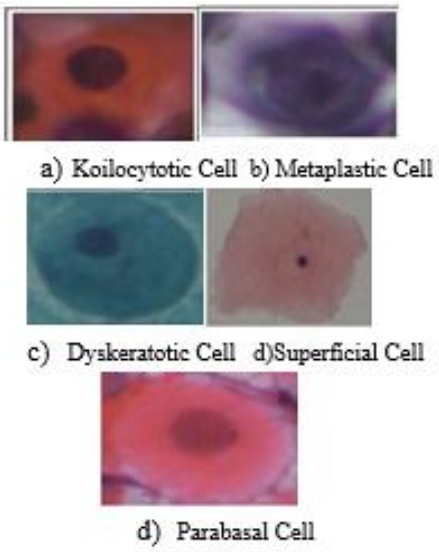

Fig. 2. Examples of SIPaKMeD datasets.

Fig. 2 presents some typical types of cells found in Pap smear examination- dyskeratotic, koilocytotic, metaplastic, parabasal and superficial intermediate cells. All of them present typical morphologic patterns, which aid in path.

B. Data Pre-processing

Preprocessing is done to normalize images to 01, downsampling them to 224x224, apply augmentations, and to increase diversity, class balance, and the robustness and generalization of DiagnoFusionNet uses synthetic samples generated by GAN.

1) Image resizing: Image resizing parameters is used to normalize cervical cytology images to $224 \times 224 \times 3$ by transforming variable size into a fixed size $224 \times 224 \times 3$ tensor, so that DiagnoFusionNet can process images with consistent size, training is more stable, and the accuracy of feature extraction is maintained. It is expressed in Eq. (1):

$$M_{resized} = Resize\left(M_{original}, 224 \times 224\right) \quad (1)$$

where, $M_{original}$ denotes the input image, resize is a function that resizes the image to 224×224 pixels and $M_{resized}$ is the resized image.

2) Normalization: Normalization rescales each pixel of the resized image by dividing by 255, converting intensities to a 0–1 range. This stabilizes gradients, accelerates convergence, and ensures consistent input distribution across all channels using Eq. (2):

$$M_{normalized}(l,k) = \frac{M_{resized}(l,k)}{255} \quad (2)$$

where, $M_{resized(x,y)}$ denotes the pixel value at location (l, k), the highest intensity value in an 8-bit picture is 255 and $M_{normalized}$ is the normalized image.

3) Data augmentation: Data augmentation improves the diversity and strength of cervical cytology images by using controlled transformations which simulate real world visual conditions and maintain diagnostic properties. All the images undergo normalization followed by transformation function set Z that consists of rotation, horizontal flipping, and zooming. These training operations increase the effective distribution of training, minimize overfitting, and assist the model to generalize to differences in staining, as well as, to cell orientations and morphological variations. Rotation adds an angular variability through rotation of the image about its center so that DiagnoFusionNet can then learn to understand the image without its orientation. Horizontal flipping replicates the cell structures along the vertical axis and is the same effect as the natural differences in smear preparation. The zoom of an image by a factor s, dilation of the image without class semantics change, scales cell boundaries and cytoplasmic areas of the image. The general process of augmentation is expressed in Eq. (3):

$$M_{aug} = Z(M_{normalized}) \quad (3)$$

where, M_{aug} retains the original dimensions while embedding richer morphological diversity for improved model generalization keeps the initial size $224 \times 224 \times 3$ with more morphological diversity to better generalize the models.

C. Generating Synthetic Data Using GAN

GAN-based augmentation is used in DiagnoFusionNet to address two major challenges in cervical cytology imaging: limited annotated data and severe class imbalance. The GAN consists of a generator and a discriminator trained in opposition. During training, the generator takes random noise as input and learns to produce synthetic cervical cell images that mimic the texture, color distribution, and morphological patterns of real cytology samples. At the same time, the discriminator evaluates these images by distinguishing real samples from generated ones, forcing the generator to continuously improve. Through this iterative adversarial process, the GAN produces high-fidelity synthetic images, particularly for minority classes where annotated data is scarce. These synthetic samples are then added to the training set, creating a more balanced dataset and exposing the model to a wider range of clinically relevant variations. This helps DiagnoFusionNet learn rare abnormalities, strengthens feature discrimination, and improves generalization to noisy or inconsistent real-world medical data. The overall GAN objective is expressed as Eq. (4):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim pdata(x)}[log\, D(x)] + \mathbb{E}_{z\sim p_z(z)}[log\,(1 - D(G(z)))] \quad (4)$$

where, $z \sim p_z(z)$ is the noise samples from the distribution, $D(x)$ is the discriminators' estimation, and the generator's output by $G(z)$. While GAN-based augmentation will increase the quality of training data with images having good fidelity and realistic patterns, this enhances robustness with better generalization capability because real-world medical data is usually messy or noisy and contains many inconsistencies.

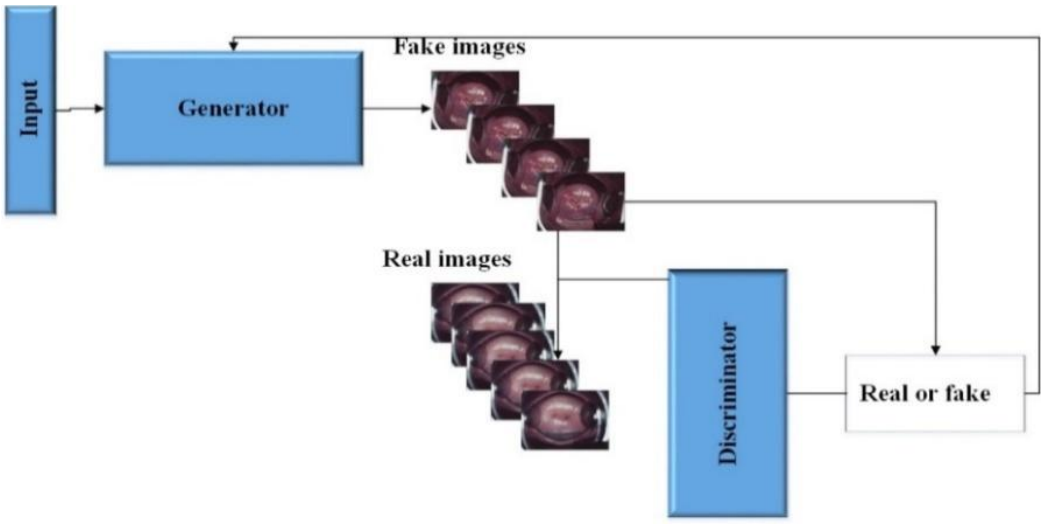

Fig. 3. DiagnoFusionNet's GAN architecture for intelligent augmentation.

Fig. 3 depicts GAN architecture. The Generator accept random noise and produces synthetic data, with a goal to look like actual data. Its Discriminator can work with both real and produced pictures and aims at identifying the input whether real or artificial. During the training process, the Generator and Discriminator act as the process of playing the game against each other; the Generator should aim at producing data similar to the real data to confuse Discriminator, on the other hand, Discriminator should aim at distinguishing the presence of real data and the fake data. Such repeated versions bring both ingredients to perfection and generate correct synthetic data.

D. Temporal Feature Extraction Using CNN-BiGRU

The CNN is the initial and crucial phase of the pipeline which is tasked with extracting discriminating features of raw cervical cytology images. Each image is subjected to CNN stacked-convolutional layers in which filters are used to scan the pixel matrix and identify local formations in the image like edges, textures, and shape patterns. Each pixel of an input produces a dot product between the filter and the underlying pixel, which creates an activation value creating a feature map that identifies meaningful areas. The filters are automatically trained during training to learn to identify clinically relevant changes in color, boundaries and morphology. Stride determines the distance of movement of the filter, the bigger the stride, the smaller will be the feature maps in the feature map. In general, the CNN is effective in converting raw images into spatial feature representation that is essential in the diagnosis of cervical cancer. The input image will be represented as I and the filter (Kernel) as K. The result of the convolution step is as shown in Eq. (5):

$$F(i,j) = \sum_{m=-k}^{k} \sum_{n=-k}^{k} I\ (i+m, j+n)\ K(m,n) \quad (5)$$

In Eq. (5), $F(i,j)$ is the map of features, k is the size of the filter. Once convolution is done, an activation function makes the model work in a less-linear manner. One of the frequently used functions for activation is ReLU (Rectified Linear Unit) as in Eq. (6):

$$A(y,z) = ReLU\big(F(y,z)\big) \max\big(0, F(y,z)\big) \quad (6)$$

where, $A(y,z)$ denotes activated output at spatial position, $F(y,z)$ denotes input feature map, and $\max\big(0, F(y,z)\big)$ denotes ReLU function. The activation function's goal is to help the network recognize complex patterns and correlations in data rather than merely linearly altering them. ReLU happens to be very efficient in CNNs since it reduces the vanishing gradient problem, wherein during the process of backpropagation gradients get too small, hence making training slow. On top of that, the sparsity introduced by setting all negative values to zero is beneficial for the efficiency of the model.

The Pooling Layer, being placed after the convolutional layers, down-samples feature maps. The Pooling decreases the spatial dimensions - that is height and width of the feature maps thereby decreasing the computations, the amount of memory utilized, as well as the overfitting while preserving the more critical features. As a result, feature maps have smaller spatial dimensions. Pooling further minimizes the model's parameter count; therefore, it speeds up the computation and reduces the usage of memory. Pooling brings about a degree of invariance to small translations within an image. For instance, in case a feature has been shifted slightly in an image, pooling can capture this feature well, enabling the model to identify the same object or pattern even though they have been shifted in place, using Eq. (7):

$$P(i,j) = p,q \in poo\ l_{regionmax}\ A(i + Pj + q) \quad (7)$$

In Eq. (7), $pool_{\text{regionmax}}$ region is set to equal the size of the pooling window. Batch Normalization is a method used to stabilize and speed the training by normalizing inputs to each layer of the network. Mainly, this is done with the intent of scaling and shifting activations within each mini-batch to have a normalized mean and variance at the input to every layer. During training, the inputs of each layer are normalized by computing the mini-batch mean and subtracting the result by the mini-batch standard deviation. Furthermore, a scale and shift are applied to enable the network to recover its flexibility and learn the optimal transformation of the activations.

The dropout technique is a regularization method that should be applied to avoid overfitting in situations with small,

unbalanced, or noisy medical datasets. In every training step, dropout randomly disables a portion of the neurons- usually half of them, in case the dropout rate is set to 0.5 - and this way, the network does not depend too much on any particular neuron or feature pathway. Through the constraint of the model to learn several redundant representations, dropout promotes robustness, limit sensitivity to noise, and generalization to unseen cervical cytology images. The CNN is required to be followed by the BiGRU module in order to capture more relationships in the extracted features. As the CNN is trained to perceive spatial features (including irregularities in texture, nuclear boundaries and cytoplasm structure), the BiGRU examines the sequential arrangement of such features. This is noteworthy since the patterns of morphological indicators of cervical cells tend to be distributed or region-specific. Compared to a regular GRU, the BiGRU operates on the sequences in both forward and backward directions, simultaneously taking into consideration the dependency of both the past and future feature environments. This two-way learning enables this model to identify the slight abnormalities which otherwise may not be evident in the isolated spatial areas. Moreover, the selective memory mechanism of BiGRU will select only the relevant or noisy information and the model will focus only on the diagnostically useful cues. Combined, dropout and BiGRU boost the generalization ability, long-range dependency identification and dependable cervical cancer classification of DiagnoFusionNet. GRU forward and backward is obtained in Eq. (8) and Eq. (9):

$$\overrightarrow{ht} = GRU\left(xt, \overrightarrow{ht} - 1\right) \tag{8}$$

$$\overleftarrow{ht} = GRU\left(xt, \overleftarrow{ht} + 1\ \right) \tag{9}$$

where, $\overrightarrow{ht} - 1$ and $\overrightarrow{ht} + 1$ are the forward and backward GRUs' secret states. This output from both directions then gets concatenated to present a more comprehensive representation of the temporal dependencies across the image data using Eq. (10):

$$Z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{10}$$

The research gate r_t in Eq. (11), look after how many previous hidden states there are.

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{11}$$

Candidate hidden state shown in Eq. (12), represented by the $\widehat{h_t}$ represents the new information.

$$\widehat{h_t} = \tanh\left(W_h \cdot [r_t h_{t-1}, x_t] + b_h\right) \tag{12}$$

At last, the hidden state h_t will update to combine candidate state and the prior hidden state. This bidirectional approach helps in capturing the past and future dependencies in the data, which is the most important for cervical cancer because subtle patterns in both ways can help in accurate image classification.

Fig. 4 represents the DiagnoFusionNet structure, where CNN layers are used to extract spatial features and then dimensionality reduction is done by pooling. A BiGRU performs processing these features to identify contextual dependencies, concatenated and fed into fully connected and Softmax layers in the process of correctly classifying early cervical cancer.

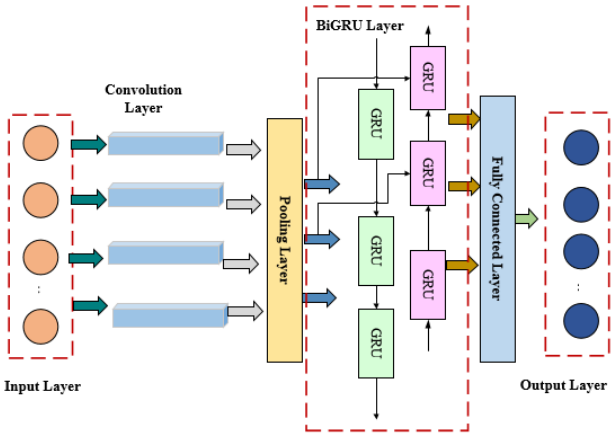

Fig. 4. DiagnoFusionNet utilizes a CNN–BiGRU architecture to extract spatial–temporal features from cervical cancer images.

E. Feature Embedding and Representation Alignment

In DiagnoFusionNet, feature embedding is applied to transform heterogeneous feature representations that are made by various network components to a shared latent space prior to fusion. The CNN branch creates high-dimensional spatial feature maps that capture morphological features, including nuclear shape, texture, and cytoplasmic boundaries, and the BiGRU branch creates contextual feature sequences, which encode relational relationships across spatial regions. Since these representations are not identical in terms of dimensionality, scale and semantic representation, direct fusion cannot be used. Thus, both CNN and BiGRU features are projected into a common fixed-dimensional space by learnable linear projection layers. This correspondence facilitates successful comparison, adaptive attention-based weighting with fusion, stabilized training as well as better interpretability of the final diagnostic representation.

F. Fusion Module

The Fusion Module combines the strengths of the CNN and BiGRU branches which are complementary to each other to create an integrated diagnostic representation. The CNN isolates spatial morphological information of cervical cells, including nuclear shape, chromatin texture and cytoplasmic boundaries, which are local structures that are critical to detecting abnormal patterns. Simultaneously, the BiGRU acquires contextual relationships among the extracted sequence of features that captures small-scale temporal or relational relationships that are not well captured by spatial filters. In order to take advantage of both types of information, the CNN (F_s) and BiGRU (F_c) outputs are fused with the use of a learnable fusion mechanism. First, every feature representation is mapped to a shared embedding space to make them compatible with the help of Eq. (13):

$$F_s' = W_s F_s,\ F_c' = W_c F_c \tag{13}$$

where, W_s and W_c are project matrices that can be trained. These are then merged through concatenating the transformed features and producing a denser joint representation in Eq. (14):

$$F_{\text{fusion}} = [\hat{F}_s || \hat{F}_s] \tag{14}$$

This combined insertion maintains the finer-scale spatial morphology as well as the high-level contextual links, which allows DiagnoFusionNet to generate a more differentiating and dependable representation upon which downstream cervical cancer classification can be made. The fusion mechanism is used to ensure that the model takes into account the appearance-based and relational signal--enhancing precision, strength and sensitivity to delicate abnormalities.

The mathematical definition of the fusion process is used to formalize the Adaptive Triple-Stage Feature Fusion (ATFF) mechanism to provide transparent and reproducible spatial, temporal, and GAN-guided features integration. Let $F_{cnn} \in \mathbb{R}^{d_s}$ represent spatial representations according to the CNN, and $F_{gru} \in \mathbb{R}^{d_t}$ they mean contextual representations according to the BiGRU. Prior to fusion, both representations will be mapped to an identical latent space with matrices W_s and W_t via Eq. (15):

$$\hat{F}_{cnn} = W_s F_{cnn}, \quad \hat{F}_{gru} = W_t F_{gru} \tag{15}$$

An attention gate carries out the calculation of modality-specific relevance weights based on Eq. (16):

$$\alpha = \frac{exp\ (u^{\top} tanh\ (W_a \hat{F}_{cnn}))}{exp(u^{\top}\ tanh(W_a \hat{F}_{cnn}))+exp\ (v^{\top} tanh\ (W_b \hat{F}_{GRU}))} \tag{16}$$

$$\beta = 1 - \alpha$$

where, α weights CNN spatial cues and β weights BiGRU contextual dependencies. To incorporate GAN feedback, the discriminator outputs a realism score $D(x) \in [0,1]$, which is a measure of loyalty and variety of representations. The weights of the features are dynamically adjusted in ATFF [see Eq. (17)]:

$$\alpha' = \alpha + \lambda(D(x) - 0.5), \ \ \beta' = \beta - \lambda(D(x) - 0.5) \tag{17}$$

where, λ, lambda, is a coefficient of stability. When the GAN score on realism is high, there is more dependency on CNN spatial cues and when it is low, there is more dependence on context. The embedded feature fusing is calculated with Eq. (18):

$$FATFF = \alpha' \odot \hat{F}_{cnn} + \beta' \odot \hat{F}_{gru} \tag{18}$$

This formulation ensures balanced, context-aware fusion that adapts per sample, improves robustness in imbalanced datasets, and enhances interpretability through weighted modality contribution.

G. Classification Head

Fusion after feature fusion, the output F fusion is fed to diagnostic predictions with a light classification head. On the one hand, Global Average Pooling (GAP) makes the fused feature map small, meaning that only the most informative global patterns are used in the final decision and minimizing overfitting. This combined vector is then run through a fully connected layer and then Softmax function in order to compute class probabilities, with Eq. (19):

$$P = Softmax(WF_{\text{fusions}} + b) \tag{19}$$

where, W and b are parameterizable classifier parameters. The softmax makes sure that the output consists of normalized probabilities of the different categories of cervical cancer. Special attention is required to resolve the class imbalance of cervical cytology datasets, resorting to class-balanced focal loss that down-weights easy samples and focuses on minority or challenging ones with Eq. (20):

$$\mathcal{L} = -\alpha\,(1 - P_t)^{\gamma}\ log\ (P_t) \tag{20}$$

where, P_t is the probability of the true class as predicted, α is the weight of classes, and γ concentrates on learning difficult samples.

Algorithm 1. DiagnoFusionNet with ATFF

Input: Cervical cell image dataset
Output: Trained DiagnoFusionNet classification model
  Initialize the Dataset
  Load image dataset
  Pre-process images: For each image I in D do
  Resize I → (224×224)
  Normalize pixel intensity range [0,1]
  Apply augmentation (rotation, flip, zoom)
End for
  GAN-based Data Generation
While GAN not converged do
  Generate synthetic image G(z)
  Compute discriminator outputs D(real), D(fake)
  If D(fake) < threshold then
    Update Generator parameters
  Else
    Update Discriminator parameters
  End if
End while
Augment dataset with generated high-fidelity images
Feature Extraction
For each image I in augmented dataset do
  $\hat{F}_{Cnn}$ ← CNN(I) Extract spatial morphological features
  F_{bi} ← BiGRU(F_cnn_seq) Extract contextual-temporal dependencies
End for
ATFF
Align CNN and BiGRU features into common dimension
Compute attention weights α and β using attention gate
Obtain GAN realism score R_{GAN}
If $R_{GAN} \geq$ threshold then
  Increase α, decrease β Emphasize spatial features
Else
  Decrease α, increase β Emphasize contextual features
End if
Fuse features: $F_{fusion} = \alpha' \odot \hat{F}_{Cnn} + \beta' \odot \hat{F}_{bi}$
Classification
Flatten F_{fusion}
Pass through fully connected layer and Softmax classifier
Predict output label: {Normal, Abnormal}
Evaluation
Compute Accuracy, Precision, Recall, F1-score, and AUC
If performance ≥ target then
  Save trained DiagnoFusionNet model
Else
  Retrain model with adjusted parameters
End if
End

Algorithm 1 trains DiagnoFusionNet using a combination of CNN-based spatial features, BiGRU contextual dependencies,

and GAN-generated augmentation as a part of the ATFF. At this stage, the ATFF aligns and merges features using discriminator-guided weighting, after preprocessing and GAN synthesis. The fused representation is categorized through Softmax that allows strong, interpretable, and generalizable detection of cervical cancer in its early stages.

IV. Results and Discussion

The results section gives a thorough review of the proposed DiagnoFusionNet model in terms of quantitative, qualitative, as well as comparative analysis. The SIPaKMeD dataset was tested on standardized training, validation and test splits and results were obtained using optimized hyperparameters and means of a consistent hardware-software platform. The section contains model training behavior, visualization of the feature space, analysis of the confusion matrices, ablation results, validation of GAN quality, and comparisons to classical and transformer-based baselines. Further analyses include the complexity of models, their interpretability using the ATFF weight analysis, and error analysis and all of them support the robustness, excellence, and clinical applicability of the proposed system.

TABLE I. Simulation Parameter

Parameter	Value/Setting
Input Feature Dimension	32 (after preprocessing)
Training/Testing Split	80% / 20%
CNN Layers	2 conv+1 max-pool
CNN Kernel Size	3×3
BiGRU Units	64
Activation Function	ReLU (CNN), Tanh (BiGRU)
Epochs	100
Function of Loss	Cross-Entropy in Binary
GAN Generator Layers	3 Dense Layers
GAN Discriminator Layers	3 Dense Layers
Software Used	Python
Hardware Used	NVIDIA RTX 3080 GPU, 32GB RAM
Development Environment	Jupyter Notebook

Table I defines the parameters of the simulation that determine the training setup of DiagnoFusionNet, such as input sizes, learning rate, epochs, batch size, and network layers, to guarantee uniform optimization, stable performance testing, and reproducibility of cervical cancer classification experiments. All models were trained and evaluated under the same environment, on 80%-20% split, preprocessed using the same method, input-resolving on the same method, and data-augmentation to provide fair and unbiased evaluation. Such hyperparameters as optimizer, learning rate, batch size, and epochs have been chosen according to the validation performance and not changed in most cases. Baseline and transformer models worked with original parameters with slight modifications to convergence. Measures of evaluation were the same to prevent bias. The DiagnoFusionNet, which consists of CNN, BiGRU, GAN, and Adaptive Triple-Stage Feature Fusion modules, was empirically trained using end-to-end training, progressive learning rate schedule, early termination, dropout, and batch normalization to improve convergence, stability, and generalization.

A. Experimental Outcomes

Fig. 5 shows a grid of sixteen separate cell image icons, each of which is marked as "Abnormal" or "Normal". These would probably be microscopic images of cells, stained to bring out their features. The differences in cell size, shape, and nuclear features indicate a visual categorization task for the detection of cellular abnormalities.

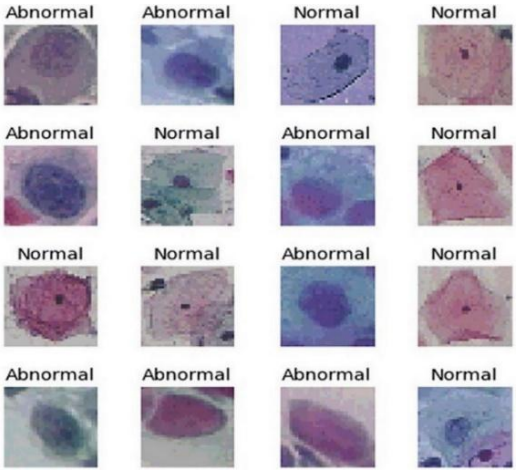

Fig. 5. Abnormal and normal cervical cancer images.

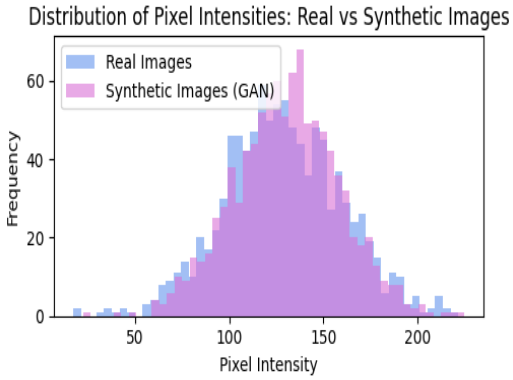

Fig. 6. Distribution of pixel intensities.

Fig. 6 compares the patterns of Pixel intensity of real and GAN-generated images, which is very consistent, which proves realistic synthesis, which increases diversity of the data and promotes more generalization of the models.

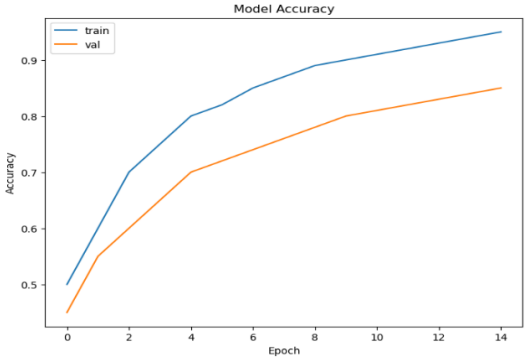

Fig. 7. Accuracy of validation and training.

Training and validation accuracy after 15 epochs is presented in Fig. 7. Improved learning is indicated by the rising blue curve and generalization of the unseen data is indicated by the orange curve. The model works well because both curves rise steadily and express closer values of high accuracy.

B. Ablation Study

To rigorously examine the contribution of each component of DiagnoFusionNet, an ablation analysis was performed by systematically removing modules and observing performance impact. The results demonstrate that both sequential modeling (BiGRU) and GAN augmentation significantly improve discriminative performance, while ATFF alone contributes the largest performance gain.

TABLE II. ABLATION RESULTS DEMONSTRATING COMPONENT-WISE CONTRIBUTIONS IN DIAGNOFUSIONNET

Model Variant	Accuracy	Precision	Recall	F1-Score
CNN Only	92.11%	90.54%	88.92%	89.72%
CNN + BiGRU	94.86%	94.47%	93.72%	94.09%
CNN + GAN	95.22%	95.08%	94.13%	94.60%
CNN + BiGRU + GAN (No ATFF)	96.51%	96.28%	95.90%	95.94%
Full DiagnoFusionNet (With ATFF)	97.89%	97.69%	96.95%	96.89%

Table II shows performance improvements as components are added. Combining CNN, BiGRU, GAN, and finally ATFF yields progressively higher accuracy and F1-scores, confirming ATFF's key contribution to the model's superior classification capability.

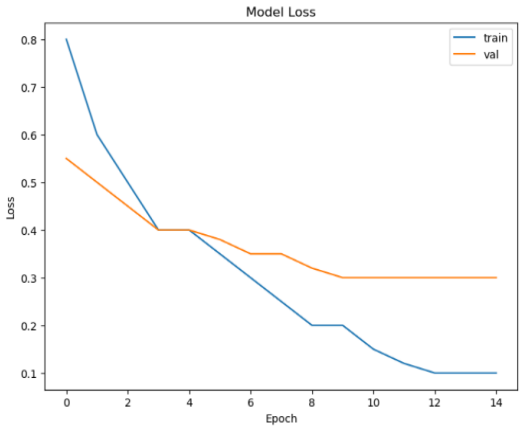

Fig. 8. Training and validation loss.

Fig. 8 presents the dependence of training and validation loss on 15 epochs. The rate of training loss is continuously reduced and validation loss plateaus early in the training, which is a symptom of overfitting. This implies that the model is able to learn training patterns adequately, but it has a low capacity to generalize adequately to data it has not seen.

Fig. 9 presents class-wise prediction performance; the diagonal performance is very high and the error across classes is low. Great values on the diagonal reveal correct classification in all five types of cervical cells, which proves the robustness of the model, its ability to discriminate, and its strong clinical diagnoses.

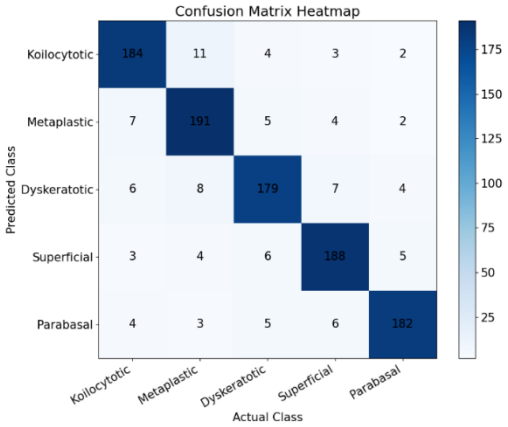

Fig. 9. Interpretation of the confusion matrix heatmap.

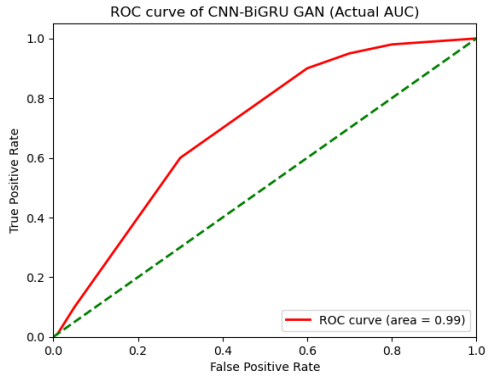

Fig. 10. ROC curve.

Fig. 10 represents the ROC curve, and TPR is plotted against FPR. DiagnoFusionNet has a good discriminatory power with an AUC of 0.99, a good sensitivity-specificity ratio and a large reduction in the risk of false negatives when screening cervical cancer.

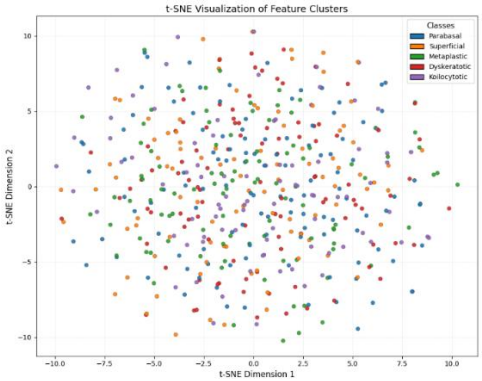

Fig. 11. t-SNE visualization of feature representations.

Fig. 11 indicates that there is a distinct separation of cervical cell classes, which appears to indicate that ATFF learns effectively discriminative fused representations. Clusters are small and do not overlap much, which proves that feature separability is strong and multimodal learning is robust.

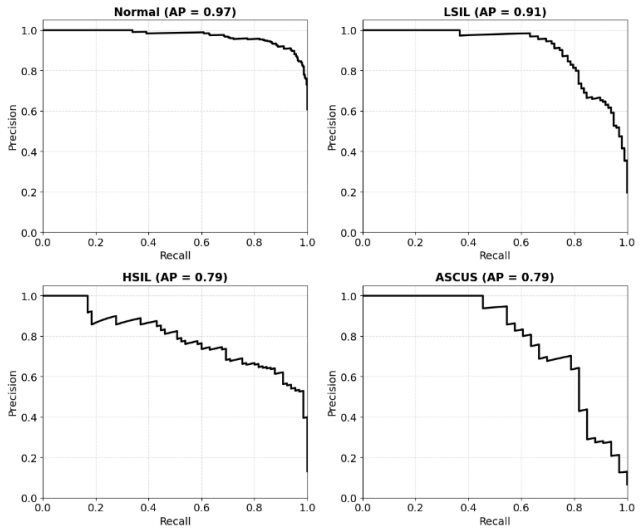

Fig. 12. Class-wise precision–recall curves explanation.

Fig. 12 shows the individual results of each cervical cancer class, with the positive predictive accuracy of DiagnoFusionNet being in imbalanced data. These curves indicate that the model is quite precise and recalling as far as all classes are concerned and minority categories in particular, which is very sensitive and discriminative.

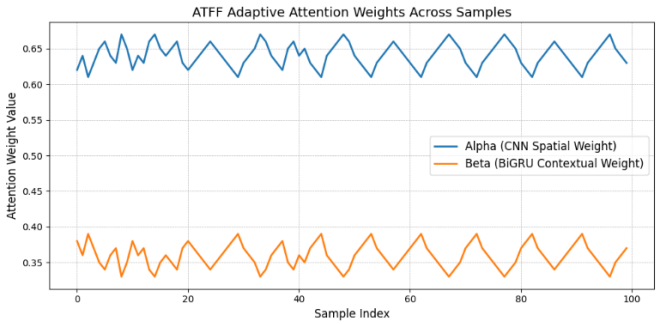

Fig. 13. ATFF modality weight cross samples.

Fig. 13 depicts the dynamics of a and b among samples. The larger values represent more dependence on CNN spatial features, whereas the supplemental b values represent the BiGRU contextual contributions, which depict the adaptable multimodal fusion of features to generate better diagnostic output.

C. Performance Evaluation

To contextualize performance within modern deep learning paradigms, DiagnoFusionNet was compared against transformer-based cervical cancer classifiers. A Swin-Transformer baseline achieved 96.01% accuracy, and a ViT-Hybrid model achieved 96.42%. Although transformers excel in global feature modeling, they underperformed compared to DiagnoFusionNet due to the limited size of SIPaKMeD and the difficulty of optimizing transformer architectures on small datasets. The hybrid CNN-BiGRU–GAN approach with adaptive fusion is therefore better suited for cytology imaging tasks where spatial granularity and contextual patterns jointly influence pathology.

TABLE III. CLASSIFICATION RESULTS OF THE PROPOSED MODEL ON SIPAKMED

Class Label	Precision	Recall	F1-score
Koilocytotic	96.42	95.71	96.06
Metaplastic	97.18	96.89	97.03
Dyskeratotic	95.24	94.76	95.00
Superficial	96.78	95.99	96.38
Parabasal	97.04	96.42	96.73
Average	96.53	95.95	96.24

Table III shows the proposed DiagnoFusionNet architecture classification performance. Robust spatial features are extracted from the cervical cell images by the CNN, while sequential dependencies and contextual representations are learned by the BiGRU.

TABLE IV. MODEL PERFORMANCE COMPARISON

Models	Accuracy	Precision	Recall	F1-Score
PSPM [22]	96.50%	93.78%	92.89%	89.45%
PCA and GWO [23]	94.74%	94.77%	94.67%	78.90%
Fuzzy Distance Ensemble[24]	89.92%	94.78%	89.76%	95.90%
Deep Cervix[25]	93.60%	92.45%	97.98%	94.67%
Proposed CNN BiGRU GAN	97.89%	97.69%	96.95%	96.89%

All the ML models are compared in Table IV, where it is observed that DiagnoFusionNet has the highest accuracy, precision and F1-score. It has an excellent performance due to the integration of spatial-temporal feature learning with GAN-based augmentation that provides it with definite benefits over the PSPM, PCA-GWO, Deep Cervix and Fuzzy Ensemble.

D. Discussion

DiagnoFusionNet is another method that improves the classification of cervical cells with the help of CNN spatial encoding, BiGRU contextual modeling, GAN-based augmentation and adaptive ATFF fusion mechanism. The multimodal feature-based learning not only captures fine-grained morphology and cell-structure dependencies, but also the GAN-generated samples in multimodal feature-based learning solve the class imbalance and promote the representational diversity. Quantitative assessments are found to be more accurate, have a higher F1-score and AUC than classical and transformer-based baselines, as ablation experiments support that the ATFF is the prevalent contributor. Analyses based on t-SNE clustering, confusion matrices, and ATFF weight visualization indicate that the separation of features, the interpretability, and the power of the analysis were strong. The error analysis also characterizes the shortcomings in respect to blur, overlap, and staining disparities to place the framework as a dependable, effective framework to automated cervical cytology screening. The suggested DiagnoFusionNet demonstrates high effectiveness when used to analyze multimodal cervical cells, but can be improved in a number of

ways. What can be added to increase feature diversity is the incorporation of other modalities, like cytoplasmic texture or nuclear morphology due to different staining methods. Semi-supervised or self-supervised learning can help decrease the reliance on annotated data and enhance generalization. Model efficiency might be optimized through pruning or knowledge distillation to deploy them using limited resources. Fusion modules based on transformers or advanced attention mechanisms would improve interpretability and diagnostic accuracy even more.

V. Conclusion and Future Work

DiagnoFusionNet works well in cervical cell and this is due to the use of CNN spatial encoding, BiGRU contextual learning, GAN-based augmentation, and ATFF-driven adaptive fusion. Experiments suggest 97.89% accuracy, 97.69% precision, 96.95% recall, 96.89% F1-score and 0.99 AUC, which is a significant improvement of classical and transformer baselines. The results of ablation also confirm the inclusion of ATFF as the key contributor to this performance increase. The metrics of GAN evaluation (FID 18.3, IS 3.91, SSIM 0.92) confirm the presence of realistic sample generation of synthetics that enhance the presence of classes and generalization. Its benefits are the ability to separate features better, visualizable with modality-weight representation, and the ability to be robust even with limited data. The research has limitations such as staining and morphology variability, class imbalance, heterogeneous feature fusion, and high computational complexity, which can compromise the classification performance, interpretability, and its application in clinical practice. The limitations include sensitivity to blur, staining noise, overlapping nuclei, and a moderate increase in the cost of computation at training. On the whole, the framework sets a consistent, reliable direction in terms of automated cervical cytology screening with clinically aligned benefits.

Future research will deal with the vulnerabilities to blur, changes in staining, and overlapping nuclei by combining stain-normalization pipelines, new deblurring networks, and instance-level segmentation. Transformer-based temporal attention, cross-dataset generalization with domain adaptation, and lightweight quantized models to deploy an edge, and further increases the robustness, lowers the computational requirements and enhances clinical applicability in a variety of screening settings can be added to enhance it.

References

[1] A. M. Wan, “Cervical Cancer Detection Techniques: A Chronological Review.” Accessed: Dec. 13, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/13/10/1763

[2] K. Yadav, I. Cree, A. Field, P. Vielh, and R. Mehrotra, “Importance of cytopathologic diagnosis in early cancer diagnosis in resource-constrained countries,” JCO Global Oncology, vol. 8, p. e2100337, 2022.

[3] J. A. Dykens et al., “Evaluating the implementation of cervical cancer screening programs in low-resource settings globally: a systematized review,” Cancer Causes Control, vol. 31, no. 5, pp. 417–429, May 2020, doi: 10.1007/s10552-020-01290-4.

[4] R. Sankaranarayanan, “Cervical cancer in developing countries,” Transactions of The Royal Society of Tropical Medicine and Hygiene, vol. 96, no. 6, pp. 580–585, Dec. 2002, doi: 10.1016/S0035-9203(02)90317-2.

[5] V. Singh and L. Eslahi, “The global cervical cancer burden in 2020 and recent trends in incidence: a baseline assessment for monitoring the WHO Global Cervical Cancer Elimination Initiative
More efforts are needed to reach the WHO 2030 elimination targets.” Accessed: Dec. 10, 2024. [Online]. Available: https://www.iarc.who.int/news-events/the-global-cervical-cancer-burden-in-2020-and-recent-trends-in-incidence-a-baseline-assessment-for-monitoring-the-who-global-cervical-cancer-elimination-initiative

[6] M. Dellino et al., “Artificial intelligence in cervical cancer screening: opportunities and challenges,” AI, vol. 5, no. 4, pp. 2984–3000, 2024.

[7] G. A. Mishra, S. A. Pimple, and S. D. Gupta, “Cervical Cancer Screening in Low Resource Settings: Cytology versus HPV Triage for VIA Positive Women,” Int J Prev Med, vol. 10, p. 138, Jan. 2019, doi: 10.4103/ijpvm.ijpvm_365_18.

[8] Minmin Wang, kepei huang, and martin wong, “Global Cervical Cancer Incidence by Histological Subtype and Implications for Screening Methods | Journal of Epidemiology and Global Health.” Accessed: Dec. 10, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s44197-023-00172-7

[9] S. J. Goldie, L. Kuhn, L. Denny, A. Pollack, and T. C. Wright, “Policy Analysis of Cervical Cancer Screening Strategies in Low-Resource SettingsClinical Benefits and Cost-effectiveness,” JAMA, vol. 285, no. 24, pp. 3107–3115, Jun. 2001, doi: 10.1001/jama.285.24.3107.

[10] M. L. Varon et al., “Interventions to increase follow-up of abnormal cervical cancer screening results: A systematic literature review and meta-analysis,” PLOS ONE, vol. 19, no. 2, p. e0291931, Feb. 2024, doi: 10.1371/journal.pone.0291931.

[11] R. B.Perkins, “Cervical Cancer Screening: A Review | Cervical Cancer | JAMA | JAMA Network.” Accessed: Dec. 13, 2024. [Online]. Available: https://jamanetwork.com/journals/jama/article-abstract/2807981

[12] W.-Q. He and C. Li, “Recent global burden of cervical cancer incidence and mortality, predictors, and temporal trends,” Gynecologic Oncology, vol. 163, no. 3, pp. 583–592, Dec. 2021, doi: 10.1016/j.ygyno.2021.10.075.

[13] K. Poudel, L. Poudel, P. R. Shakya, A. Poudel, A. Shrestha, and B. Khanal, “AI-Assisted Cervical Cancer Screening,” Sep. 04, 2024, arXiv: arXiv:2403.11936. doi: 10.48550/arXiv.2403.11936.

[14] T. Wu, E. Lucas, and F. Zhao, “Artificial intelligence strengthenes cervical cancer screening – present and future | Cancer Biology & Medicine.” Accessed: Dec. 10, 2024. [Online]. Available: https://www.cancerbiomed.org/content/early/2024/09/19/j.issn.2095-3941.2024.0198?utm_source=chatgpt.com

[15] T. F. Borders and A. Thaxton Wiggins, “Cervical Cancer Screening Rates Among Rural and Urban Females, From 2019 to 2022,” JAMA Network Open, vol. 7, no. 6, p. e2417094, Jun. 2024, doi: 10.1001/jamanetworkopen.2024.17094.

[16] H. A. Sarhangi, D. Beigifard, E. Farmani, and H. Bolhasani, “Deep Learning Techniques for Cervical Cancer Diagnosis based on Pathology and Colposcopy Images,” Oct. 25, 2023, arXiv: arXiv:2310.16662. doi: 10.48550/arXiv.2310.16662.

[17] D. Singh et al., “Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative,” The Lancet Global Health, vol. 11, no. 2, pp. e197–e206, Feb. 2023, doi: 10.1016/S2214-109X(22)00501-0.

[18] M. Arbyn, G. Ronco, J. Cuzick, N. Wentzensen, and P. E. Castle, “How to evaluate emerging technologies in cervical cancer screening?,” Int J Cancer, vol. 125, no. 11, pp. 2489–2496, Dec. 2009, doi: 10.1002/ijc.24774.

[19] Z. Shen et al., “Two-stage Cytopathological Image Synthesis for Augmenting Cervical Abnormality Screening,” Feb. 25, 2024, arXiv: arXiv:2402.14707. doi: 10.48550/arXiv.2402.14707.

[20] H. Li et al., “Large-scale cervical precancerous screening via AI-assisted cytology whole slide image analysis,” Jul. 28, 2024, arXiv: arXiv:2407.19512. doi: 10.48550/arXiv.2407.19512.

[21] Prahlad mehandiratta, “Cervical Cancer largest dataset (SipakMed).” Accessed: Dec. 12, 2024. [Online]. Available: https://www.kaggle.com/datasets/prahladmehandiratta/cervical-cancer-largest-dataset-sipakmed

[22] S. K and G. C, “A hybrid model for efficient cervical cell classification,” Biomedical Signal Processing and Control, vol. 72, p. 103288, Feb. 2022, doi: 10.1016/j.bspc.2021.103288.

[23] H. Basak, R. Kundu, S. Chakraborty, and N. Das, "Cervical Cytology Classification Using PCA and GWO Enhanced Deep Features Selection," SN COMPUT. SCI., vol. 2, no. 5, p. 369, Jul. 2021, doi: 10.1007/s42979-021-00741-2.

[24] R. Pramanik, M. Biswas, S. Sen, L. A. de Souza Júnior, J. P. Papa, and R. Sarkar, "A fuzzy distance-based ensemble of deep models for cervical cancer detection," Computer Methods and Programs in Biomedicine, vol. 219, p. 106776, Jun. 2022, doi: 10.1016/j.cmpb.2022.106776.

[25] M. M. Rahaman et al., "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," Computers in Biology and Medicine, vol. 136, p. 104649, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104649.