Deep Learning-based Method in Multimodal Data for Diabetic Retinopathy Detection

Kartina Diah Kesuma Wardhani [a,b], Shahreen Kasim [a,*], Aldo Erianda [c], Rohayanti Hassan [d]

[a] Soft Computing and Data Mining Centre (SMC), Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Parit Raja, Johor, Malaysia
[b] Department of Informatics Engineering, Politeknik Caltex Riau, Pekanbaru, Indonesia
[c] Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia
[d] Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
Corresponding author: *shahreen@uthm.edu.my

Abstract— Diabetic retinopathy (DR) is a complex condition, and incorporating information from multiple sources, such as patient history, laboratory results, or genetic data, can enhance understanding. An ophthalmologist or an automated system can identify DR through manual examination. The automatic detection of diabetic retinopathy has become a preferred choice for patients and healthcare providers due to its cost-effectiveness and time efficiency. The novelty of this research lies in developing a model for predicting diabetic retinopathy using multimodal data fusion, incorporating fundus retinal images, optical coherence tomography (OCT), and electronic health records (EHR) through an early fusion technique implemented in a Long Short-Term Memory (LSTM) network. Our model, which utilizes an early fusion of multimodal data with Local Binary Pattern (LBP), has demonstrated the best performance, achieving an AUC value of 0.99. This high accuracy indicates that integrating information from various data sources can significantly improve the capability of the model in detecting both positive and negative cases of diabetic retinopathy, instilling confidence in the reliability of our findings.

Keywords— Diabetic retinopathy; multimodal; LSTM; LBP.

I. INTRODUCTION

Various conditions and diseases can affect the retina, leading to vision problems. Examples include macular degeneration, diabetic retinopathy (DR), and retinal detachment [1]. Diabetic retinopathy (DR) arises as a complication of diabetes, impacting the blood vessels within the light-sensitive tissue located at the posterior part of the eye. It is a prevalent complication associated with diabetes and is a primary contributor to adult blindness [2]–[4] .The identification of DR can be accomplished through manual examination by an ophthalmologist or by utilizing an automated system [2], [4]. The ophthalmologist will begin by taking a detailed medical history, namely an electronic health record (EHR), including information about the patient's diabetes, overall health, and any existing eye conditions [5]. Subsequently, physicians interpret the features within retinal images to categorize cases of DR [6]. The impact of manual interpretation relies on the clinician's level of expertise and experience [7]. The automatic detection of diabetic retinopathy (DR) has become a preferred choice for patients and healthcare providers due to its cost-effectiveness and time efficiency [8].

Typically, DR detection using machine learning involves retinal image processing utilizing fundus image analysis [3], [9]–[11]. These images encompass the posterior segment of the eye, including the retina, optic disc, macula, and blood vessels. Additionally, 3D optical coherence tomography (OCT) and 3D OCT angiography have been employed for DR detection [12]–[15]. It captures volumetric data, facilitating a more comprehensive evaluation of retinal structures. Machine learning is also used for data analysis of electronic health records (EHRs) in DR detection [5]. DR detection methods are referred to as unimodal learning or single-modality learning. However, utilizing single-modality data for DR detection presents challenges: Diabetic retinopathy is a complex condition, and incorporating information from multiple sources, such as patient history, laboratory results, or

genetic data, can enhance understanding. Unimodal learning may overlook capturing these multimodal relationships.

This research aims to improve the diagnostic accuracy of diabetic retinopathy (DR) by fusing retinal imaging datasets and electronic health records (EHR). Specifically, the retinal imaging used in this study includes fundus retinal images and optical coherence tomography (OCT). This study also employs EHR to predict DR. The aim is to identify the relationship between these three modalities in DR prediction. Multimodal learning is performed by fusing these three data modalities using a deep learning approach. Recent studies have applied convolutional neural networks (CNN) for image data modeling [7], [16]–[18], long short-term memory (LSTM) networks to address the vanishing or exploding gradient effect [19], [20], deep neural networks (DNN) for tasks such as classification and regression on static data like images, text, or other structured data [21], [22], and temporal neural networks (TNN) to handle temporal relationships in data, enabling the model to understand and leverage patterns in sequential data.

This research models the prediction of diabetic retinopathy using multimodal data fusion of fundus retinal images, optical coherence tomography, and electronic health records, implemented with a fusion technique in LSTM networks.

II. MATERIALS AND METHOD

We utilized an open dataset of OCT from https://borealisdata.ca/dataset [23], fundus retinal images from an open dataset by Kaggle [24], and EHR from the open dataset by Dr. Yad. The OCT [25] data consists of 98 NoDR patients and 49 DR patients. The fundus dataset consists of 245 DR cases and 286 NoDR cases. The EHR dataset comprises 426 entries, including 260 patients with diabetes mellitus without DR and 168 patients with diabetes mellitus with DR, with 31 features. For prediction, we assigned the label 0 to patients with DR and 1 to patients without DR. Figure 1 illustrates examples of normal eye conditions and DR conditions in fundus images and OCT images.

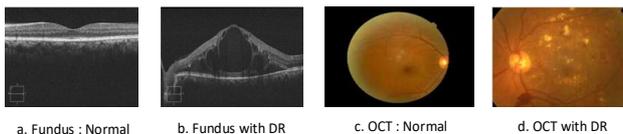a. Fundus : Normal   b. Fundus with DR   c. OCT : Normal   d. OCT with DR

Fig. 1 Example of pairs of Fundus Retinal Images and OCT Retinal Images

A. Preprocessing Data

The research process involved several scenarios, as depicted in Figure 2. Data preprocessing was conducted through feature extraction on the image datasets, including fundus and OCT images. Feature extraction in this study is crucial because diabetic retinopathy (DR) is an eye condition caused by diabetes that can damage the retina's blood vessels, which can be observed through fundus or OCT images. In the fundus and OCT datasets, the texture of the retinal images can provide essential information about the structure and condition of the retina [26], [27]Several texture features may be important in detecting DR through fundus and OCT images, including microaneurysms, exudates, abnormal blood vessels, and structural damage.

Feature extraction in this study employs the Local Binary Pattern (LBP) technique due to its efficacy in extracting texture features from images [28]–[34]. LBP works according to the following flows [35], [36]:

a. Define a neighborhood: Select a pixel in the image as the central pixel and define a circular neighborhood around it. The most common neighborhood consists of 8 surrounding pixels.

b. Thresholding: Compare the intensity value of each neighbor with the intensity value of the central pixel. If the neighbor's intensity is greater than or equal to the central pixel's intensity, assign a value of 1; otherwise, assign a value of 0.

c. Binary Pattern: Arrange the binary values in a clockwise or counterclockwise order to form a binary number.

d. Decimal Conversion: Convert the binary number to a decimal value. This value represents the LBP code for the central pixel.

e. Histogram Generation: Repeat the above steps for all pixels in the image and generate a histogram of the LBP codes. This histogram serves as the image's texture descriptor.

The feature extraction output is a feature vector, which is then concatenated with the EHR dataset. In our implementation, we augmented all modalities to enhance the diversity of the training data by creating variations of existing samples without introducing new labels. This was done to prevent overfitting and improve the model's generalization capabilities [17],[37]–[40]. Augmentation was also performed to standardize the dimensions of each modality to 2000. The fused features will subsequently serve as the input data for the modeling phase, ultimately leading to the output of DR (Diabetic Retinopathy) detection.
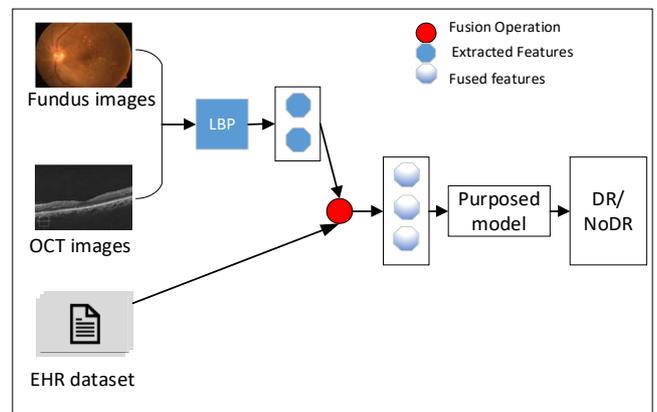

Fig. 2 Proposed Pipeline Multimodal Data Fusion for DR Detection

B. Proposed DL Model

The fusion of multimodal data in the previous stage resulted in a fused dataset that will be modeled using deep learning (DL) for DR detection. In this study, the DL approach is Long Short-Term Memory (LSTM) with multimodal data input.

The fusion of multimodal data in the previous stage resulted in a fused dataset that will be modeled using deep learning (DL) for DR detection. In this study, the DL approach used is Long Short-Term Memory (LSTM) with multimodal data input. LSTM consists of cells, each with three main gates: input gate, forget gate, and output gate. At

each time step t, the LSTM performs several operations to update the cell state and hidden state [41], [42]:

   a. Forget Gate Activation:

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \qquad (1)$$

Where W_f is the forget gate weight, $[h_{t-1}, x_t]$ is the concatenation of the previous hidden state and the current input, and b_f is the bias.

   b. Input Gate Activation:

$$i_t = \sigma(W_f . [h_{t-1}, x_t] + b_i) \qquad (2)$$

Where W_f is the input gate weight and b_i the bias.

   c. Candidate Cell State:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \qquad (3)$$

Where Wc is the weight for updating the cell state and bc is the bias.

   d. Cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \qquad (3)$$

The cell state is updated by combining the previous cell state filtered by the forget gate and the candidate cell state filtered by the input gate.

   e. Output Gate Activation:

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \qquad (4)$$

Where W_o is the output gate weight and b_o is the bias.

   f. Hidden State Update:

$$h_t = o_t * \tanh(C_t) \qquad (5)$$

The hidden state is updated using the value from the output gate and the transformed cell state by the $tanh$ function.

The LSTM architecture, as shown in Figure 3, consists of several main layers that work together to process the input and generate the output. Here is a detailed explanation of each layer in the architecture; Input Layer consists of data from various modalities that have been fused. This data can be in text, images, or similar data combined into a single representation; LSTM Layer 1 comprises 50 LSTM units. LSTM (Long Short-Term Memory) is a neural network that handles sequential data by maintaining contextual information over long periods. This layer processes the input from the previous layer and generates output that will be passed to the next layer. The dropout layer prevents overfitting by setting a dropout rate of 20%. This means that during training, 20% of the units in this layer will be randomly deactivated for each forward pass, ensuring that the model does not become overly reliant on specific units and remains more generalized. LSTM Layer 2 also consists of 50 LSTM units, similar to the first LSTM layer. This layer processes the output from the dropout layer and sends the result to the next layer. Dense Layer Output has two output neurons with a SoftMax activation function. These two output neurons represent two classes: DR (e.g., disease detection) and NoDR (no disease detection). The output Layer consists of two labels DR and NoDR. This architecture is designed to process sequential data from various modalities through LSTM layers and then use dropout to prevent overfitting and classification.

III. RESULT AND DISCUSSION

The model developed in this research was implemented using the Keras library on an NVIDIA GeForce RTX 3080. One of the key metrics used for model evaluation and classification is the confusion matrix, which is divided into four parts describing the predicted results against the actual results. The confusion matrix consists of four terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [43], [44].

These metrics include accuracy, precision, recall (or sensitivity), specificity, and Area Under the Curve (AUC). These measurements provide a comprehensive understanding of the model's performance in different aspects of classification.
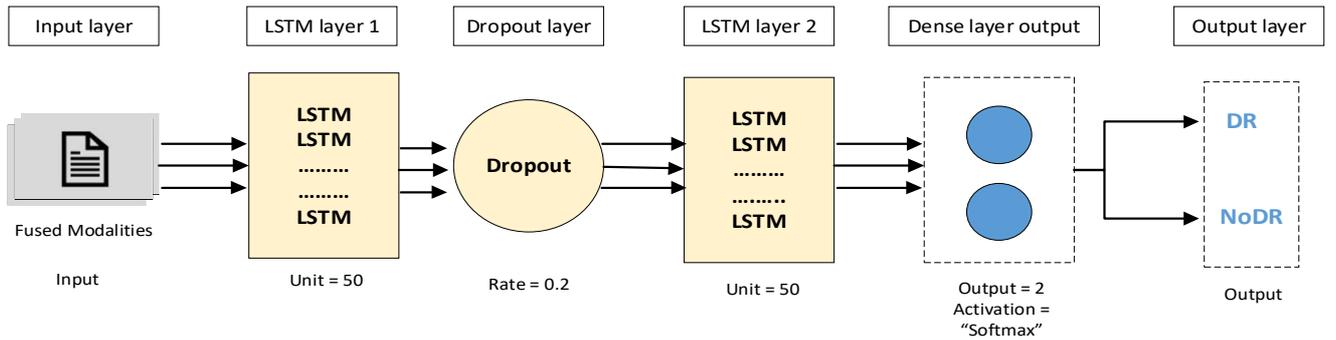
Fig. 3 Proposed Model

A. Comparison of LSTM-Based Method with Unimodal Data

Figures 4, 5, and 6 depict the training and validation during unimodal images fundus, OCT, and EHR modeling using LSTM. Figure 4 shows the training and validation loss chat of Fundus Images with LBP using LSTM; at the start of training, the training and validation data loss values dropped significantly, indicating that the model learned well. After several epochs, the loss values in the training and validation data continue to decrease and stabilize, indicating that the model has reached convergence. There is no clear indication of overfitting because the validation loss value does not increase significantly. In the training and validation accuracy chart, at the beginning of training, the accuracy values on the training and validation data increase rapidly, indicating that the model learns quickly. After several epochs, the accuracy values on the training and validation data are stable and high, indicating that the model has good generalization ability. Accuracy values on training and validation data are relatively stable at around 85%

to 90%, indicating consistent performance. Overall, the model shows outstanding performance with low loss values and high accuracy on training and validation data. The model has good generalization ability, with validation accuracy almost equivalent to training accuracy.

Figure 5 shows the training and validation loss chart of OCT Images with LBP using LSTM; at the start of training, the training and validation data loss values dropped significantly, indicating that the model learned well. After several epochs, the training and validation data loss values are stable and low, indicating that the model has reached convergence. At the end of the training, there was a slight increase in the validation loss value, which could indicate overfitting to the training data. At the training and validation accuracy chart, at the beginning of training, the accuracy values on the training and validation data increase rapidly, indicating that the model learns quickly. After several epochs, the accuracy values on the training and validation data are stable and high, indicating that the model has good generalization ability. The accuracy value of the training data is slightly higher than the validation data, but both remain stable at around 95% to 100%. Overall, the model performs excellently with low loss values and high accuracy on training and validation data. The model has good generalization ability, with validation accuracy almost equivalent to training accuracy.

Figure 6 consists of two graphs that show changes in loss and accuracy values during training and validation of the LSTM model on the Unimodal EHR (Electronic Health Records) dataset. The chart above consists of two graphs that show changes in loss and accuracy values during training and validation of the LSTM model on the Unimodal EHR (Electronic Health Records) dataset. In the training and validation loss chart, at the start of training, the loss values on the training and validation data dropped significantly, indicating that the model was learning well. After around 20 epochs, the loss values on the training and validation data

continue to decrease, with the validation loss value being lower than the training loss value.

In the training and validation accuracy chart, the accuracy values on the training and validation data increase at the beginning of training, indicating that the model learns quickly. The validation accuracy value reaches around 70% after several epochs and remains constant without fluctuation, indicating stability in the validation data. Overall, the model shows a sound reduction in loss and increased accuracy at the beginning of training. Still, there is a significant difference between the training and validation accuracy values. After several epochs, the model reaches stability at a validation accuracy value of around 70%. This graph shows that the LSTM model performs well on the Unimodal EHR dataset initially but may suffer from overfitting and stability issues on the training data.

Figure 7 illustrates the training loss and accuracy during modeling multimodal images using LSTM. Training and Validation Loss chart at the start of training, the loss values on the training and validation data dropped significantly, indicating that the model learned well. After about 20 epochs, the training and validation data loss values are stable and low, indicating that the model has reached convergence. There are no clear indications of overfitting because the validation loss and training loss values remain balanced. Training and Validation Accuracy chart: at the beginning of training, the accuracy values on the training and validation data increase rapidly, indicating that the model learns quickly. After about 20 epochs, the accuracy values on the training and validation data are stable and high, indicating that the model has good generalization ability. Accuracy values on training and validation data are relatively stable at around 90% to 100%, indicating consistent performance. Overall, the model performs well with low loss values and high accuracy on training and validation data. The model has good generalization ability, with validation accuracy almost equivalent to training accuracy.
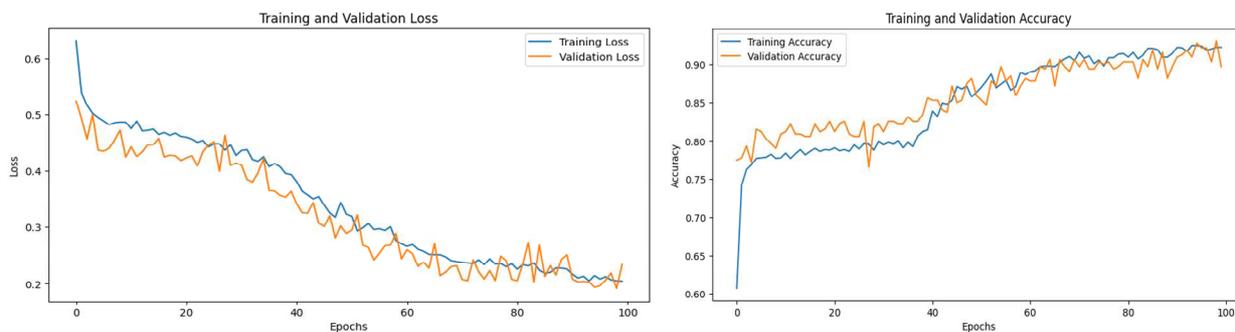

Fig. 4 Training and validation chart Fundus Images with LBP Using LSTM
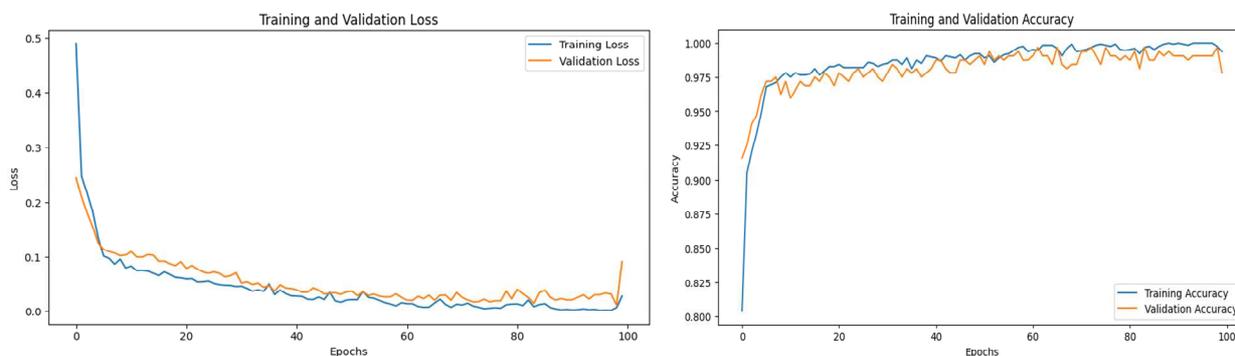

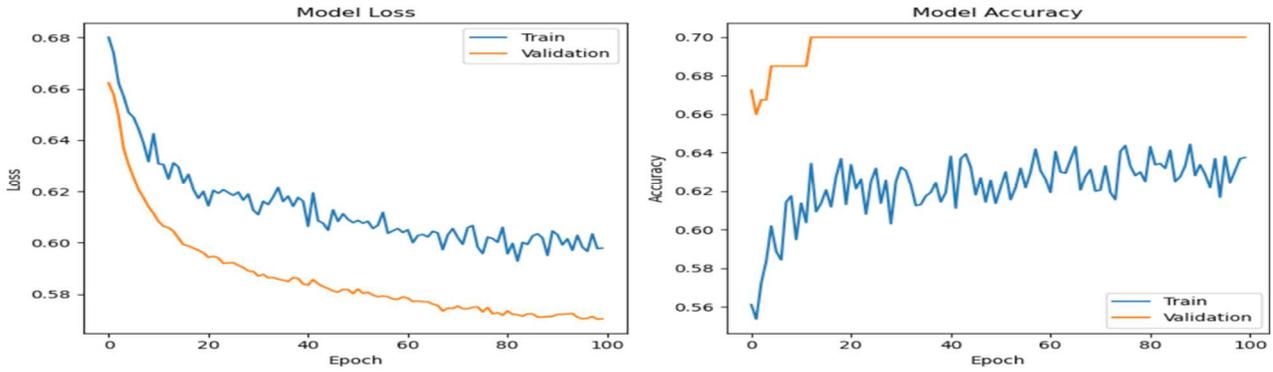Fig. 5 Training and validation chart OCT Images with LBP Using LSTM

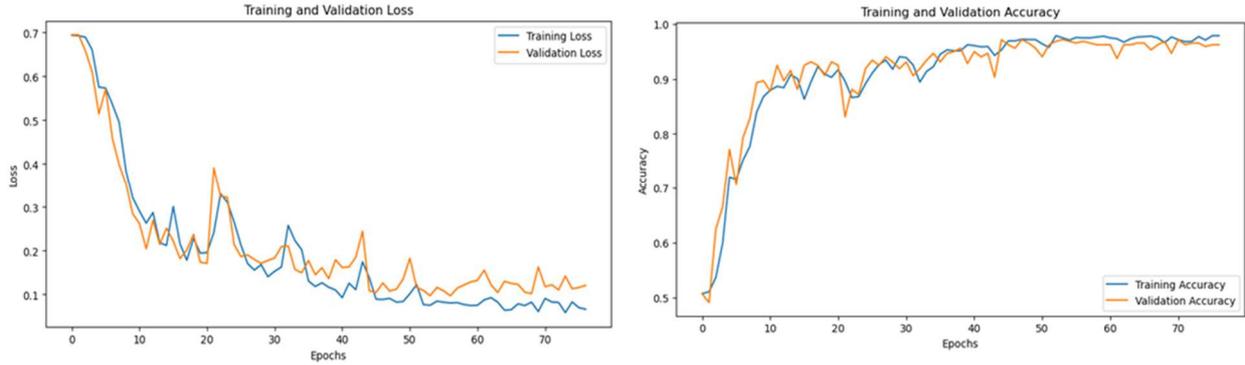Fig. 6 Training and Validation chart EHR using LSTM

Fig. 7 Multimodal Fusion training and validation Chart with LBP Using LSTM

The comparison of unimodal images with multimodal images conducted in this study is illustrated in Figure 8. Unimodal Fundus + LBP performed well but lacked sensitivity compared to other methods with an AUC value of 0.96. Unimodal OCT + LBP show excellent performance with a high sensitivity of 0.98, accuracy, and near-perfect AUC. Unimodal EHR showed lower performance than image-based methods, with the lowest sensitivity. Early Fusion with LBP performed best overall, especially for an AUC nearly perfect at 0.99. Combining information from multiple data sources provides excellent results.

The multimodal data with the LBP method provides the best results in all evaluation metrics, indicating that combining data from various sources can improve model performance. Unimodal OCT + LBP also showed excellent performance, especially regarding sensitivity and AUC. Unimodal Fundus + LBP was good in accuracy and specificity but slightly inferior in sensitivity. Unimodal EHR

performed the lowest among all methods, suggesting that image-based data may be more informative for this task than EHR data alone. Combining data from various sources with fusion multimodal data and LBP techniques for feature extraction is very effective in improving model performance in detecting health conditions.

B. Comparison with Other Existing Multimodal Data Fusion Research in DR

We compared the performance of multimodal data fusion based on LSTM with similar studies that used different modalities and methods. Figure 9 illustrates this comparison, which shows the AUC (Area Under the Curve) values from various multimodal fusion studies alongside the current research. AUC measures a model's ability to distinguish between positive and negative classes, with a value of 1.0 indicating perfect performance.

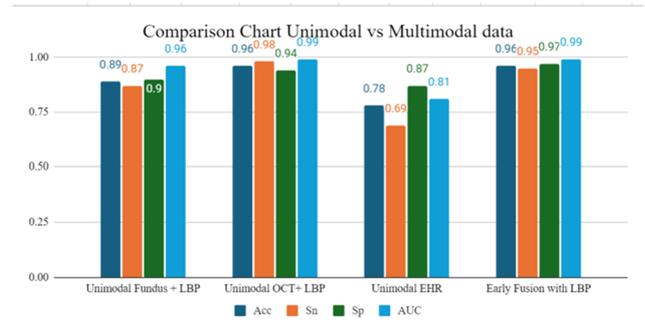|	Acc	Sn	Sp	AUC
Unimodal Fundus + LBP	0.89	0.87	0.9	0.96
Unimodal OCT+ LBP	0.96	0.98	0.94	0.99
Unimodal EHR	0.78	0.69	0.87	0.81
Multimodal + LBP	0.96	0.95	0.97	0.99

Fig. 8 Performance comparison unimodal vs multimodal data in DR Detection

Table 1 compares various research studies on applying multimodal data fusion for diabetic retinopathy detection. It includes information on the publication year, datasets used, methods employed, fusion strategies, and each study's Area Under the Curve (AUC) performance metric. In 2023, Y. Li et al. utilized the Evired dataset, which includes HR-OCTA and UWF-OCTA images. They applied deep learning backbones such as ResNet, DenseNet, and EfficientNet, implementing early and late fusion strategies, achieving an AUC of 0.87. Hsu et al., in 2022, used fundus images and EHR information, combining CNN with Multilayer Perceptron (MLP) in an early fusion strategy, resulting in an AUC of 0.97. Similarly, El et al. 2023 worked with UWF-CFP and OCTA datasets, employing CNN models (ResNet50 and 3D-ResNet50), and achieved an AUC of 0.8175 with early fusion.

In 2020, V. Tseng et al. used private fundus images and the Messidor-2 dataset, applying CNN models (Inception-v4 and DenseNet) in a two-stage fusion process involving both late fusion and early fusion strategies, achieving an AUC of 0.97 using Messidor-2. The current research conducted in 2024 utilized fundus images, OCT images, and EHR data, combining Local Binary Pattern (LBP) with Long Short-Term Memory (LSTM) in an early fusion strategy, achieving the highest AUC of 0.99. This table underscores the variety of datasets, methods, and fusion strategies used across different studies, highlighting the continuous advancements and diverse approaches in improving the accuracy of diabetic retinopathy detection through multimodal data fusion. The highest AUC achieved by the current research demonstrates the effectiveness of combining LBP and LSTM in an early fusion strategy.
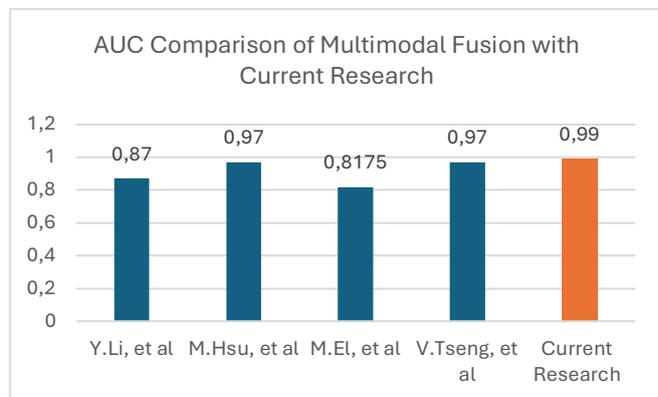

Fig. 9 AUC Comparison Chart

TABLE I
THE EXISTING METHOD OF MULTIMODAL FUSION IN DR

Research	Pub Year	Dataset	Method	Fusion Strategy	AUC
Li et al. [45]	2023	Evired: HR-OCTA, UWF-OCTA	DL Backbone: ResNet, DenseNet and EfficientNet	Early Fusion and Late Fusion	AUC: 0.87
Hsu et al [46]	2022	Fundus Images & EHR Information	CNN + MLP (Multilayer Perceptron)	Early Fusion	AUC: 0.97
El et al. [47]	2023	UWF-CFP and OCTA	CNN (ResNet50 and 3D-ResNet50).	Early Fusion	AUC: 0.8175
Tseng et al [48]	2020	Private Fundus images and Messidor-2	CNN (Inception-v4 and DensNet) Two-stages Fusion: CNN (RetinaNet, Inception-v4)	Late Fusion and Early Fusion	AUC Using Messidor-2: 0.97
Current Research	2024	Fundus Images, OCT Images & EHR Dataset	LBP + LSTM	Early Fusion	AUC: 0.99

IV. CONCLUSIONS

This research demonstrates that the proposed model for DR detection using multimodal data fusion is a highly effective approach for enhancing the model's ability to distinguish between positive and negative classes. Integrating information from various data sources can significantly improve the accuracy and capability of the model in detecting both positive and negative cases. The ongoing research development aims to optimize data preprocessing to reduce the impact of the curse of dimensionality from multimodal fusion. Further applications of the developed model consider the potential to achieve DR detection with more optimal performance and processing time, such as developing modeling using multimodal fusion with joint fusion and late fusion techniques. This will allow the resulting model to make more optimal predictions and gain broader utility in DR detection.

REFERENCES

[1] N. Marchesi, F. Fahmideh, F. Boschi, A. Pascale, and A. Barbieri, "Ocular Neurodegenerative Diseases: Interconnection between Retina and Cortical Areas," *Cells*, vol. 10, no. 9, p. 2394, Sep. 2021, doi:10.3390/cells10092394.

[2] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Informatics in Medicine Unlocked*, vol. 20, p. 100377, 2020, doi:10.1016/j.imu.2020.100377.

[3] I. Qureshi, J. Ma, and Q. Abbas, "Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11691–11721, Jan. 2021, doi: 10.1007/s11042-020-10238-4.

[4] N. Mukherjee and S. Sengupta, "Application of deep learning approaches for classification of diabetic retinopathy stages from fundus retinal images: a survey," *Multimedia Tools and Applications*, vol. 83, no. 14, pp. 43115–43175, Oct. 2023, doi: 10.1007/s11042-023-17254-0.

[5] W.-C. Lin, J. S. Chen, M. F. Chiang, and M. R. Hribar, "Applications of Artificial Intelligence to Electronic Health Record Data in Ophthalmology," *Translational Vision Science & Technology*, vol. 9, no. 2, p. 13, Feb. 2020, doi: 10.1167/tvst.9.2.13.

[6] D. A. Hasan, S. R. M. Zeebaree, M. A. M. Sadeeq, H. M. Shukur, R. R. Zebari, and A. H. Alkhayyat, "Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems- A Survey," *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*, vol. 4, pp. 16–21, Apr. 2021, doi:10.1109/bicits51482.2021.9509920.

[7] A. Sungheetha and R. Sharma R, "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network," *Journal of Trends in Computer Science and Smart Technology*, vol. 3, no. 2, pp. 81–94, Jul. 2021, doi:10.36548/jtcsst.2021.2.002.

[8] A. Skouta, A. Elmoufidi, S. Jai-Andaloussi, and O. Ouchetto, "Deep learning for diabetic retinopathy assessments: a literature review,"

Multimedia Tools and Applications, vol. 82, no. 27, pp. 41701–41766, Apr. 2023, doi: 10.1007/s11042-023-15110-9.

[9] K. Parthiban and M. Kamarasan, "Diabetic retinopathy detection and grading of retinal fundus images using coyote optimization algorithm with deep learning," *Multimedia Tools and Applications*, vol. 82, no. 12, pp. 18947–18966, Nov. 2022, doi: 10.1007/s11042-022-14234-8.

[10] S. Kannoju, "Diabetic Retinopathy Detection Using Retinal Images," Master's thesis, California State University, Sacramento, 2021.

[11] A. M. Dayana and W. R. S. Emmanuel, "A Comprehensive Review of Diabetic Retinopathy Detection and Grading Based on Deep Learning and Metaheuristic Optimization Techniques," *Archives of Computational Methods in Engineering*, vol. 30, no. 7, pp. 4565–4599, Jun. 2023, doi: 10.1007/s11831-023-09946-5.

[12] N. K. Waheed et al., "Optical coherence tomography angiography in diabetic retinopathy," *Progress in Retinal and Eye Research*, vol. 97, p. 101206, Nov. 2023, doi: 10.1016/j.preteyeres.2023.101206.

[13] M. M. Khansari et al., "Automated Deformation-Based Analysis of 3D Optical Coherence Tomography in Diabetic Retinopathy," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 236–245, Jan. 2020, doi: 10.1109/tmi.2019.2924452.

[14] D. Gildea, "The diagnostic value of optical coherence tomography angiography in diabetic retinopathy: a systematic review," *International Ophthalmology*, vol. 39, no. 10, pp. 2413–2433, Oct. 2018, doi: 10.1007/s10792-018-1034-8.

[15] J. Chua et al., "Optical Coherence Tomography Angiography in Diabetes and Diabetic Retinopathy," *Journal of Clinical Medicine*, vol. 9, no. 6, p. 1723, Jun. 2020, doi: 10.3390/jcm9061723.

[16] A. M. R. Godla, D. Kavitha, K. Ashok, M. S. A. Ansari, N. Satheesh, and R. V. K. Reddy, "An Ensemble Learning Approach for Multi-Modal Medical Image Fusion using Deep Convolutional Neural Networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023, doi: 10.14569/ijacsa.2023.0140884.

[17] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprevost, "Reducing Overfitting and Improving Generalization in Training Convolutional Neural Network (CNN) under Limited Sample Sizes in Image Recognition," *2020 - 5th International Conference on Information Technology (InCIT)*, Oct. 2020, doi:10.1109/incit50588.2020.9310787.

[18] V. Vaidyaa and L.K. Vishwamitra, "Diabetes Detection using Convolutional Neural Network through Feature Sequencing", *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, pp. 2783-2789, 2021.

[19] H. Zhang, H. Huang, and H. Han, "Attention-Based Convolution Skip Bidirectional Long Short-Term Memory Network for Speech Emotion Recognition," *IEEE Access*, vol. 9, pp. 5332–5342, 2021, doi:10.1109/access.2020.3047395.

[20] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, May 2020, doi: 10.1007/s10462-020-09838-1.

[21] C. Schröder and A. Niekler, "A survey of active learning for text classification using deep neural networks", *arXiv*:2008.07267, 2020.

[22] G. Mena, K. Coussement, K. W. De Bock, A. De Caigny, and S. Lessmann, "Exploiting time-varying RFM measures for customer churn prediction with deep neural networks," *Annals of Operations Research*, vol. 339, no. 1–2, pp. 765–787, Mar. 2023, doi:10.1007/s10479-023-05259-9.

[23] P. Gholami, P. Roy, and V. Lakshminarayanan, 'Diabetic Retinopathy Retinal OCT Images'. Borealis, 2018.

[24] P. Darabi, 'Diagnosis of Diabetic Retinopathy'. 15-Jul-2024, doi: 10.13140/RG.2.2.13037.19688.

[25] X. Zhuang et al., "Data from: Association of diabetic retinopathy and diabetic macular oedema with renal function in southern Chinese patients with type 2 diabetes mellitus: a single-centre observational study," [Dataset]. *Dryad*, Sep. 2019, doi:10.5061/dryad.6kg1sd7.

[26] M. Monemian, M. Irajpour, and H. Rabbani, "A review on texture-based methods for anomaly detection in retinal optical coherence tomography images," *Optik*, vol. 288, p. 171165, Oct. 2023, doi:10.1016/j.ijleo.2023.171165.

[27] K. Zhou et al., "Encoding Structure-Texture Relation with P-Net for Anomaly Detection in Retinal Images," *Computer Vision – ECCV 2020*, pp. 360–377, 2020, doi: 10.1007/978-3-030-58565-5_22.

[28] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1311–1328, Jun. 2020, doi: 10.1007/s00521-020-05017-z.

[29] K. Kaplan, Y. Kaya, M. Kuncan, M. R. Minaz, and H. M. Ertunç, "An improved feature extraction method using texture analysis with LBP

[30] S. Hamad Khaleefah, S. A. Mostafa, A. Mustapha, and M. Faidzul Nasrudin, "Review of local binary pattern operators in image feature extraction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, p. 23, Jul. 2020, doi:10.11591/ijeecs.v19.i1.pp23-31.

[31] N. Kaur, N. Nazir, and Manik, "A Review of Local Binary Pattern Based texture feature extraction," *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, vol. 475, pp. 1–4, Sep. 2021, doi:10.1109/icrito51393.2021.9596485.

[32] Y. Luo, J. Sa, Y. Song, H. Jiang, C. Zhang, and Z. Zhang, "Texture classification combining improved local binary pattern and threshold segmentation," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 25899–25916, Mar. 2023, doi: 10.1007/s11042-023-14749-8.

[33] M. H. Shakoor, R. Boostani, M. Sabeti, and M. Mohammadi, "Feature selection and mapping of local binary pattern for texture classification," *Multimedia Tools and Applications*, vol. 82, no. 5, pp. 7639–7676, Aug. 2022, doi: 10.1007/s11042-022-13470-2.

[34] R. Tekin, Ö. F. Ertuğrul, and Y. Kaya, "New local binary pattern approaches based on color channels in texture classification," *Multimedia Tools and Applications*, vol. 79, no. 43–44, pp. 32541–32561, Aug. 2020, doi: 10.1007/s11042-020-09698-5.

[35] A. El Idrissi, Y. El merabet, and Y. Ruichek, "Palmprint recognition using state‐of‐the‐art local texture descriptors: a comparative study," *IET Biometrics*, vol. 9, no. 4, pp. 143–153, May 2020, doi:10.1049/iet-bmt.2019.0103.

[36] T. Chen et al., "A novel face recognition method based on fusion of LBP and HOG," *IET Image Processing*, vol. 15, no. 14, pp. 3559–3572, Apr. 2021, doi: 10.1049/ipr2.12192.

[37] R. Balestriero and L. Bottou, "The Effects of Regularization and Data Augmentation are Class Dependent," *Neural Information Processing Systems (NeurIPS)*, 2022.

[38] E.H. Hssayni, M. Ettaouil, "Generalization Ability Augmentation and Regularization of Deep Convolutional Neural Networks Using L1/2 Pooling", *International Journal on Technical and Physical Problems of Engineering (IJTPE)*, Issue 48, Vol. 13, No. 3, pp. 1-6, September 2021.

[39] C. Khosla and B. S. Saini, "Enhancing Performance of Deep Learning Models with different Data Augmentation Techniques: A Survey," *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, Jun. 2020, doi:10.1109/iciem48762.2020.9160048.

[40] H. Yao et al., 'Improving Generalization in Meta-learning via Task Augmentation', in *Proceedings of the 38th International Conference on Machine Learning*, 18--24 Jul 2021, vol. 139, pp. 11887–11897.

[41] P. Zhang, C. Li, C. Peng, and J. Tian, "Ultra-Short-Term Prediction of Wind Power Based on Error Following Forget Gate-Based Long Short-Term Memory," *Energies*, vol. 13, no. 20, p. 5400, Oct. 2020, doi: 10.3390/en13205400.

[42] J. Oruh, S. Viriri, and A. Adegun, "Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition," *IEEE Access*, vol. 10, pp. 30069–30079, 2022, doi: 10.1109/access.2022.3159339.

[43] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi:10.1109/access.2022.3151048.

[44] I. M. De Diego, A. R. Redondo, R. R. Fernández, J. Navarro, and J. M. Moguerza, "General Performance Score for classification problems," *Applied Intelligence*, vol. 52, no. 10, pp. 12049–12063, Jan. 2022, doi:10.1007/s10489-021-03041-7.

[45] Y. Li et al., "Hybrid Fusion of High-Resolution and Ultra-Widefield OCTA Acquisitions for the Automatic Diagnosis of Diabetic Retinopathy," *Diagnostics*, vol. 13, no. 17, p. 2770, Aug. 2023, doi:10.3390/diagnostics13172770.

[46] M.-Y. Hsu et al., "Deep Learning for Automated Diabetic Retinopathy Screening Fused With Heterogeneous Data From EHRs Can Lead to Earlier Referral Decisions," *Translational Vision Science & Technology*, vol. 10, no. 9, p. 18, Aug. 2021, doi: 10.1167/tvst.10.9.18.

[47] M. El Habib Daho et al., "Improved Automatic Diabetic Retinopathy Severity Classification Using Deep Multimodal Fusion of UWF-CFP and OCTA Images," *Ophthalmic Medical Image Analysis*, pp. 11–20, 2023, doi: 10.1007/978-3-031-44013-7_2.

[48] V. S. Tseng et al., "Leveraging Multimodal Deep Learning Architecture with Retina Lesion Information to Detect Diabetic Retinopathy," *Translational Vision Science & Technology*, vol. 9, no. 2, p. 41, Jul. 2020, doi: 10.1167/tvst.9.2.41.

for bearing fault diagnosis," *Applied Soft Computing*, vol. 87, p. 106019, Feb. 2020, doi: 10.1016/j.asoc.2019.106019.