International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

# A DEEP LEARNING MULTI-MODAL FEATURE FUSION BASED DISEASE CLASSIFICATION WITH SMOTE RE-SAMPLING TECHNIQUE

**[1]Dr. B. Sateesh Kumar, [2]Mummadi Swathi**

[1] Professor & Head of Department, Department of Computer Science and Engineering, JNTUH College of Engineering, Jagtial, Telangana, India.

[2] M. Tech, Department of CSE, JNTUH College of Engineering, Jagtial, Telangana, India.
mummadi.swathi6@gmail.com

**Abstract:** The exploration presents a deep learning-based multi-modal feature fusion disease classification model. Consolidating chest X-ray pictures and ailment data further develops classification accuracy. A novel versatile multi-modal attention strategy progressively consolidates feature vectors from the two faculties to support grouping execution. The openI Chest X-ray dataset approves the model. Test extension utilizing the SMOTE strategy addresses overfitting, low recall, and F1 score brought about by little and lopsided examples. Removal concentrates on show that multi-modular models beat single-modular models that utilize just pictures or text. The versatile multi-modal consideration method beats vector connection for feature fusion. Results show that the proposed approach decreases test awkwardness and further develops sickness order accuracy, expanding recall and F1 scores. This study shows that multi-modal data and refined combination approaches can reinforce clinical indicative DL models.

**Index Terms:** Diseases classification, multi-modal feature fusion, self-attention, SMOTE

## 1. INTRODUCTION

With the rapid advancement and widespread adoption of Artificial Intelligence (AI) technologies, research in Machine Learning (ML) and Deep Learning (DL) algorithms, which drive cutting-edge AI innovations, has become increasingly significant. One of the prominent applications of AI in healthcare is the development of intelligent medical auxiliary systems designed for disease classification. These systems utilize a variety of multi-modal features, including patient lesion images, textual descriptions of symptoms, and structured data on relevant physical indicators, to train and optimize classifiers that assist doctors in making accurate diagnoses [1].

Historically, traditional statistical methods have been employed for disease classification both domestically and internationally. These methods have provided a solid foundation for medical diagnostics, leveraging statistical inference to analyze patient data and predict disease outcomes. However, with the emergence and progression of DL technologies, there has been a notable shift towards using deep neural networks for disease classification. These networks have demonstrated remarkable success, significantly improving

the accuracy and reliability of medical diagnoses [2].

Modern ML theory often incorporates regression techniques within its framework, treating them as a subset of ML methods. These techniques, along with various ML classifiers, are integral to disease classification tasks. Commonly used ML classifiers include Support Vector Machines (SVMs), Conditional Random Fields (CRFs), and Random Forests (RFs). For instance, SVMs have been employed to identify cancer patients by analyzing gene expression data, and when combined with Recursive Feature Elimination (RFE), they have shown improved accuracy compared to conventional classification methods [3]. SVM and Multi-Layer Perceptron (MLP) have been used for classifier based features fusion [5].

The fusion of multi-modal data—combining different types of information such as images, text, and structured data—has become a critical approach in enhancing the performance of disease classification models. This approach leverages the complementary strengths of various data types, providing a more comprehensive representation of patient health status. For example, chest X-ray images can reveal structural abnormalities in the lungs, while textual descriptions of symptoms can provide context and detail that may not be apparent in the images alone [8].

In this study, proposed a disease classification model based on multi-modal feature fusion using DL techniques. Our model integrates chest X-ray images and corresponding disease descriptions, aiming to leverage the rich information present in both visual and textual data. A key innovation of our approach is the introduction of an adaptive multi-modal attention mechanism, which dynamically fuses feature vectors extracted from both modalities. This mechanism allows the model to focus on the most relevant features from each modality, enhancing the overall classification performance [9].

To validate our model, the Chest X-ray dataset from the openI database is used. Given the challenges posed by small and imbalanced datasets, which can lead to overfitting and low recall and F1 scores, the Synthetic Minority Over-sampling Technique (SMOTE) is used to expand the sample size and balance the dataset.

Our experiments include an ablation study to compare the performance of our multi-modal model with single-modal models using only images or text, as well as with a model using simple vector concatenation for feature fusion [6]. The multimodal system has been designed at multiclassifier & multimodal level. At multi-classifier level, multiple algorithms are combined gives better results.

The results of our study demonstrate that the proposed multi-modal model significantly outperforms single-modal models, achieving higher classification accuracy, recall, and F1 scores. The adaptive multi-modal attention mechanism also shows superior performance compared to simple vector concatenation, indicating its effectiveness in enhancing feature fusion. These findings underscore

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

the potential of combining multi-modal data and advanced fusion techniques to improve the robustness and effectiveness of DL models in medical diagnostics [7].

## 2. PROBLEM STATEMENT

The ongoing framework utilizes BiLSTM-CRF model to separate clinical text features from radiological reports and customary ML classifiers likeSVM, RF, DT, and Logistic Regression to predict liver cancer.

In another paper, they made FAMLC-BERT, which utilizes consideration component and is significant in NLP for multi-label illness text categorization. Their methodology utilizes BERT multi-label classification to gather semantic features at various levels and component level thoughtfulness regarding perceive electronic clinical record labels.

Overfitting can result from include level consideration. A model that overemphasizes specific features during training may not sum up well to new information, bringing about lackluster showing on new examples. Setting mindfulness can be obliged by highlight level consideration processes that attention on information elements or channels.

Past endeavors didn't utilize resampling procedures like SMOTE to address overfitting, low recall, and F1 esteem attributable to little and imbalanced data.

## 3. FRAMEWORK OF PROPOSED WORK

A DL-based disease classification model utilizing multi-modal feature fusion is proposed. It involves chest X-ray pictures and sickness data as picture and text modal data. Utilize a versatile multi-modal attention strategy to blend these changed information sorts.

This strategy powerfully combines feature vectors from the two modalities. With the openI Chest X-ray dataset, the recommended model is shown effective.

SMOTE addresses overfitting, low recall, and F1 values brought about by little and uneven examples. SMOTE balances picture and text modes by extending the dataset.

This strategy prepares the model on a more delegate dataset, helping speculation and versatility. Multi-modular information, versatile consideration components, and SMOTE example development further develop characterization exactness, making ailment detection more accurate.

## 4. SYSTEM DESIGN

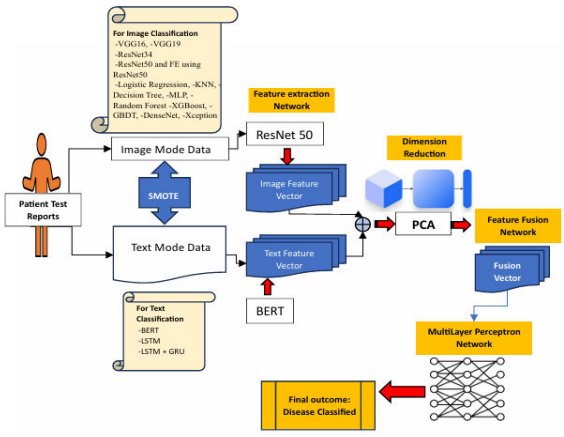


**Fig.4.1.** System Architecture

The proposed multi-modal feature fusion illness classification system architecture has a few basic parts. Subsequent to gathering and preprocessing chest X-ray pictures and sickness portrayals, picture and text modular information are made. Convolutional neural networks (CNNs) remove qualities from chest X-ray pictures, though RNNs or transformers handle text based portrayals.

Versatile multi-modal attention is then used to meld the removed feature vectors from the two modalities progressively. This keeps the classifier zeroed in on the main information type properties. The SMOTE extends the dataset and balances tests across all modalities to lessen overfitting and information awkwardness.

The fused feature vectors are provided into a completely associated neural network classifier to group diseases. OpenI's Chest X-ray dataset approves the framework's viability.

**c) Dataset Collection:**

This study involved the openI information base of labeled clinical pictures for its chest X-ray dataset. This assortment contains assortment of chest X-ray pictures with sickness analyze and clinical comments. The dataset contains a few pneumonic sicknesses, making it ideal for infection order model preparation and approval. The information gathering technique contains downloading photographs and text information for designing and consistency. An assortment of photographs and composed portrayals is made by marking each picture with its clinical conclusion. The multi-modal feature fusion model is created and tried on this arranged dataset, empowering solid ailment order.

**d) Image Processing:**

This review requires numerous significant image processing cycles to create chest X-ray pictures for feature extraction and classification. X-ray pictures are preprocessed to work on quality and consistency. To keep up with dataset consistency, photographs are resized to 224x224 pixels. Contrast change and standardization further develop picture clearness and predictable pixel esteem circulations for convolutional neural network feature extraction.

Then, pivot, flipping, and scaling are utilized to misleadingly grow the dataset and make changeability to diminish overfitting. Further developing the model's speculation requires this step. Sound decrease advances like Gaussian obscuring can wipe out unessential subtleties and stress critical X-ray attributes.

In the wake of handling the photos, a CNN removes progressive components including edges, surfaces, and structures vital to sickness location utilizing numerous convolutional layers. Picture methodology information and text data are utilized in the multi-modular consideration system for sickness order. This broad picture handling pipeline gives excellent information to exact and vigorous model execution.

**e) Algorithms:**

Various arrangements of calculations, like those for image classification (- VGG16, -

VGG19, - ResNet34, - ResNet50), and FE using ResNet50 - Logistic Regression, are utilized to group the text mode and picture mode information. Decision Tree, -KNN, - Random Forest and MLP - XGBoost, - GBDT, - DenseNet, - Xception, and - BERT - LSTM - LSTM + GRU for text classification. For this situation, the information is re-examined utilizing SMOTE.

The SMOTE strategy is utilized to expand the example size in both the text and picture modes. It can likewise resolve issues with little and imbalanced example measures that lead to overfitting, low recall, and F1 esteem.

## 5. ALGORITHMS FOR MULTI-MODAL FEATURE FUSION

## FOR IMAGE CLASSIFICATION

A set of computer operations called image classification algorithms analyzes image data to determine the label or category that most accurately represents the image. Several image classification methods are used in this model, including:

**i) VGG16:** Oxford Visual Geometry Group developed a 16-layer convolutional neural network architecture, VGG16. It uses deep layers and 3x3 convolution filters to extract hierarchical features from images. Thanks to VGG16's deep architecture and powerful feature learning capabilities, it achieves high accuracy in image classification tasks.

**ii) VGG19:** By adding three additional layers to VGG16, VGG19 now has a total of 19 layers. Keeping the same deep architecture and 3x3 convolution filter size improves the model's ability to capture complex patterns and details in images. VGG19 is widely used in computer vision applications to accurately classify images due to its reliable performance.

**iii) ResNet34:** ResNet34 is an adaptation of the Residual Network (ResNet) architecture that contains 34 layers (i.e., iii). The vanishing gradient problem in deep networks is solved by introducing residual blocks with skip connections. ResNet34 efficiently learns hierarchical features from photos, making it suitable for tasks that require detailed feature extraction and classification at low computational cost.

**iv) ResNet50:** ResNet50 adds 50 layers and more residual blocks to ResNet34 to enable deeper feature learning. Skip connections are still used to improve the efficiency of model training and gradient flow. ResNet50 is known for its excellent performance in image classification tasks. By leveraging its deeper architecture and feature extraction capabilities, it achieves state-of-the-art accuracy on benchmark datasets.

**v) Feature Extraction (FE) using ResNet50:** FE uses a pre-trained ResNet50 to extract high-level features from images. These features can be used later for tasks such as detection and classification without having to retrain the entire network.

**vi) Logistic Regression:** A linear model for binary classification, logistic regression is used. It uses a logistic function to calculate the probability and can be applied to analyze how features affect classification decisions in an image collection.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

**vii) KNN (K-Nearest Neighbors):** KNN is a non-parametric approach that can be applied in regression and classification. It labels data points according to the majority labels of its nearest neighbors in the feature space, making it suitable for image classification problems with clear feature boundaries.

**viii) Decision Tree:** To create a tree structure for classification, decision trees repeatedly divide the data into subgroups according to characteristics. It captures nonlinear correlations in visual data in an easy-to-understand and interpretable way for classification based on learned rules.

**ix) MLP (Multi-Layer Perceptron):** Multilayer Perceptron or MLP is a type of feed-forward neural network with multiple hidden layers. It uses forward and back propagation to learn complex patterns in image data, making it well suited for tasks that require deep feature learning and classification.

**x) Random Forest:** To increase accuracy and robustness for classification problems, Random Forest is an ensemble approach that combines many decision trees. Combining predictions from multiple trees efficiently handles high-dimensional image data.

**xi) XGBoost (Extreme Gradient Boosting):** This gradient boosting technique is known for its effectiveness and speed. It generates sequential trees while optimizing a differentiable loss function, making it useful for image classification tasks where high accuracy and efficiency are essential.

**xii) GBDT (Gradient Boosting Decision Tree):** By correcting errors in previous models, GBDT creates an ensemble of decision trees to improve performance. It is very good at identifying complex visual patterns and achieving high classification accuracy.

**xiii) DenseNet:** DenseNet uses feedforward connections to connect each layer to every other layer. In deep image classification networks, it improves gradient flow and mitigates the vanishing gradient problem by enhancing feature propagation and promoting feature reuse.

**xiv) Xception:** Xception is a deep CNN architecture designed to outperform the traditional Inception module. It focuses on depth-separable convolutions to efficiently capture the spatial hierarchy in images and maximize computation and model performance.

## TEXT CLASSIFICATION ALGORITHMS

A text classification algorithm is a set of mathematical operations that analyzes the content of a text and assigns the correct label or category. By processing and analyzing the text, the algorithm finds features and patterns that distinguish one category from another. Some of the text classification algorithms used in this model are:

**i) BERT (Bidirectional Encoder Representations from Transformers):** A transformer based model pre-trained on a large text corpus. BERT stands for Bidirectional Encoder Representation from Transformers. It is capable of capturing the bidirectional context of text data, thus

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

achieving superior accuracy and contextual understanding to fine-tune various text classification tasks.

**ii) LSTM (Long Short-Term Memory):** A type of RNN that can detect long-term dependencies in sequential data thanks to its memory cells. It is suitable for text classification problems that require incremental understanding of the context.

**iii) LSTM+GRU (Gated Recurrent Unit):** This technique combines GRU and LSTM units to exploit their mutual advantages in managing sequential data. It improves the performance of text classification models by efficiently capturing complex relationships and remote context.

**Application of SMOTE:**

**Step-1:** Identify Minority Samples

$X_{min} = \{x_i | y_i = y_{min}\}$

where ymin is the label for the minority class.

**Step-2:** Set the Amount of Oversampling

Let NNN be the number of s ynthetic samples to generate.

Let kkk be the number of nearest neighbors.

**Step-3**: Find Nearest Neighbors

For each sample $x_i$ ? $X_{min}$:

$NNk(xi) = \{x_{i,1}, x_{i,2}, \ldots, x_{i,k}\}$

**Step-4:** Generate Synthetic Samples

For each sample $x_i$ ? $X_{min}$:

For j=1 to ?N/|$X_{min}$|?

Randomly select one of its k-nearest neighbors, $x_{i,nn}$ ? $NN_k(x_i)$.

**Step-5:** Generate a synthetic sample

$x_{new} = x_i + \delta^*(x_{i,nn} - x_i)$, where $\delta \sim U(0,1)$ is a random number uniformly distributed between 0 & 1.

For example, Assume a minority dataset with points {(1,1),(2,2)} and majority dataset with points {(5,5),(6,6),(7,7),(8,8)}.

**1.Identify Minority Samples:**

Xmin={(1,1),(2,2)}

**2.Set the Amount of Oversampling:**

Desired number of synthetic samples N = 2.

Number of nearest neighbors k = 1.

**3. Find Nearest Neighbors:**

For (1,1), nearest neighbor is (2,2).

For (2,2), nearest neighbor is (1,1).

**4.Generate Synthetic Samples:**

**For (1,1):**

Select neighbor (2,2).

Random number δ=0.5.

New sample:

(1,1)+0.5*((2,2)- (1,1)) = (1,1)+0.5*(1,1) = (1.5,1.5)

**For (2,2):**

Select neighbor (1,1), Random number δ=0.7.

New sample:

(2,2)+0.7*((1,1)- (2,2)) = (2,2)+0.7*(- 1,- 1) = (2- 0.7,2- 0.7) = (1.3,1.3)

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Thus, after adding synthetic samples to the minority class, the final dataset would be:

Minority class: {(1,1),(2,2),(1.5,1.5),(1.3,1.3).

Majority class: {(5,5),(6,6),(7,7),(8,8).

## 4. FINDINGS FROM THE EXPERIMENT

Following the resampling of the information utilizing SMOTE, the classification calculations for Logistic, KNN, DT, RF, XGB, GBC, MLP, and Voting showed eminent varieties in Accuracy, Precision, Recall, and F1score for every calculation.

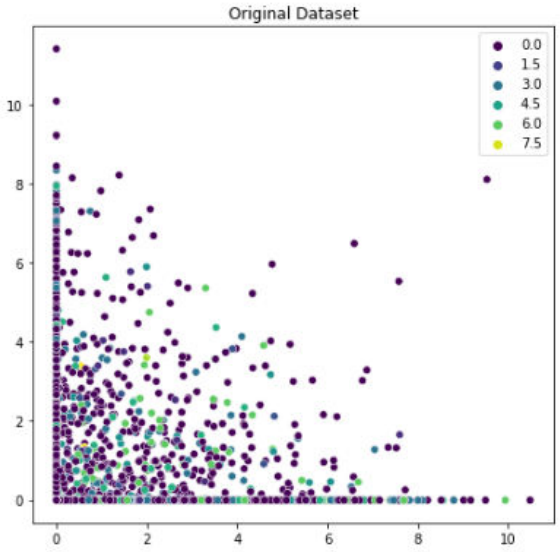


**Fig.4.1. D**ataset before applying SMOTE.

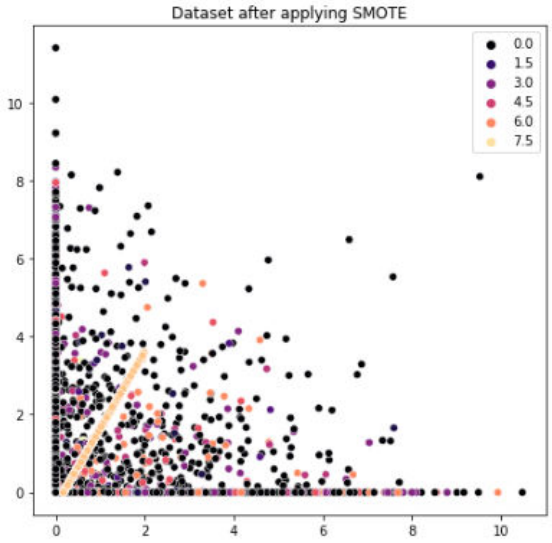


**Fig.4.2.** Dataset after applying SMOTE.

**Evaluation Metrics**: Metrics including accuracy, precision, recall, and F1-score are used to evaluate the model's performance:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

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

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where the terms true positives, true negatives, false positives, and false negatives, respectively, are TP, TN, FP, and FN.

The following table displays the performance metrics for each algorithm based on recall, accuracy, precision, and F1score:

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

| | Algorithm | Accuracy | Precision | Recall | F1score |
|---|---|---|---|---|---|
| 0 | Logistic | 0.800889 | 0.708620 | 0.800889 | 0.744230 |
| 1 | KNN | 0.784889 | 0.760835 | 0.784889 | 0.768681 |
| 2 | DT | 0.739556 | 0.734576 | 0.739556 | 0.736721 |
| 3 | RF | 0.812444 | 0.683867 | 0.812444 | 0.736162 |
| 5 | XGB | 0.810667 | 0.726803 | 0.810667 | 0.755378 |
| 6 | GBC | 0.796444 | 0.716164 | 0.796444 | 0.743337 |
| 8 | MLP | 0.734222 | 0.747452 | 0.734222 | 0.740132 |
| 9 | voting | 0.813333 | 0.721590 | 0.813333 | 0.741744 |

**Fig.4.2. A**ccuracy, Precision, Recall and F1score metrics of Logistic, KNN, DT, RF, XGB, GBC, MLP, Voting algorithms after application of SMOTE.

**KNN:** K-Nearest Neighbors
**DT:** Decision Tree
**RF:** Random Forest
**XGB:** Extreme Gradient Boosting
**GBC:** Gradient-Boosting
**MLP:** Multi-layer Perceptron
**SMOTE:** Synthetic Minority Over-sampling Technique

**Statistical Significance Testing**: To compare the performance of the proposed model with the baseline model, hypothesis testing can be used. Methods such as t-test and Wilcoxon signed rank test can be used to check the significance of the results.

**Dimensionality Reduction and Feature Selection:**

**PCA/ICA**: The feature space can be reduced by applying dimensionality reduction techniques such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA).

$$Z = XW$$

where, **W** projects the original feature space **X** on **Z** to a lower dimension.

## 5. SUMMARY OF FINDINGS

All in all, this examination gives a robust illness classification model that utilizes imaginative multi-modular element combination to beat restricted and unequal datasets. The program further develops classification accuracy over single-modular procedures by consolidating chest X-ray pictures and ailment depictions utilizing versatile multi-modal attention. The SMOTE technique lessens overfitting, low recall, and F1 score, boosting model trustworthiness and speculation.

The concentrate likewise shows that versatile multi-modal attention beats vector link, accentuating the meaning of effectively answering key data from every methodology for viable disease order. The discoveries affirm that image and text modalities cooperate to figure out clinical issues.

Future review will explore dimensionality decrease approaches past Principal Component Analysis (PCA) to expand feature extraction and computing effectiveness without compromising classification accuracy. Development in model plan and feature engineering means to make AI-driven disease arrangement systems more useful in clinical settings, further developing healthcare results and navigation.

## 6. FUTURE SCOPE

This disease classification model might develop to utilize multi-modal deep learning structures. Examination might zero in on hybrid models that consolidate hereditary information, physiological signs, pictures,

and text. High level attention components or transformer-based plans could help the model catch muddled multi-modal dataset interdependencies. Extra data augmentation and versatile examining methods could increment model strength and speculation, empowering more precise and adaptable healthcare symptomatic devices.

## REFERENCES

[1] I. Guyon, J. Weston, S. Barnhill, and V. VapnikGene, ‘‘Selection for cancer classification using support vector machines,’’ Mach. Learn., vol. 46, nos. 1–3, pp. 389–422, Jan. 2002.

[2] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, ‘‘Machine learning applications in cancer prognosis and prediction,’’ Comput. Struct. Biotechnol. J., vol. 13, pp. 8–17, Jan. 2015.

[3] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, ‘‘Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,’’ IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, Feb. 2016.

[4] Z. Tariq, S. K. Shah, and Y. Lee, ‘‘Lung disease classification using deep convolutional neural network,’’ in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), San Diego, CA, USA, Nov. 2019, pp. 732–735.

[5] B. Sateesh. Kumar and A. Govardhan, "An Efficient Boosting Approach for Score Level Fusion of Face and Palmprint Biometrics in Human Recognition," International Journal of Science and Research (IJSR), vol. 4, no. 10, Oct. 2015.

[6] H. Liu, Y. Xu, Z. Zhang, N. Wang, Y. Huang, Y. Hu, Z. Yang, R. Jiang, and H. Chen, ‘‘A natural language processing pipeline of Chinese freetext radiology reports for liver cancer diagnosis,’’ IEEE Access, vol. 8, pp. 159110–159119, 2020.

[7] D. Pan, X. Zheng, W. Liu, M. Li, M. Ma, Y. Zhou, L. Yang, and P. Wang, ‘‘Multi-label classification for clinical text with feature-level attention,’’ in Proc. IEEE 6th Int. Conf. Big Data Secur. Cloud (BigDataSecurity), IEEE Int. Conf. High Perform. Smart Comput. (HPSC), IEEE Int. Conf. Intell. Data Secur. (IDS), Baltimore, MD, USA, May 2020, pp. 186–191.

[8] J. Y. Choi, T. K. Yoo, J. G. Seo, J. Kwak, T. T. Um, and T. H. Rim, ‘‘Multicategorical deep learning neural network to classify retinal images: A pilot study employing small database,’’ PLoS ONE, vol. 12, no. 11, pp. 1–16, Nov. 2017.

[9] X. Li, H. Wang, H. He, J. Du, J. Chen, and J. Wu, ‘‘Intelligent diagnosis with Chinese electronic medical records based on convolutional neural networks,’’ BMC Bioinf., vol. 20, no. 1, pp. 1–12, Feb. 2019.

[10] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, ‘‘COVID-19 detection through transfer learning using multimodal imaging data,’’ IEEE Access, vol. 8, pp. 149808–149824, 2020.

[11] Z. Tariq, S. K. Shah, and Y. Lee, ‘‘Multimodal lung disease classification using deep convolutional neural network,’’ in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Seoul, Korea (South), Dec. 2020, pp. 2530–2537.

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ‘‘BERT: Pre-training of deep bidirectional transformers for language understanding,’’ 2018, arXiv:1810.04805.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556

[15] M. Martinc, F. Haider, S. Pollak, and S. Luz, "Temporal integration of text transcripts and acoustic features for Alzheimer's diagnosis based on spontaneous speech," Frontiers Aging Neurosci., vol. 13, pp. 1–15, Jun. 2021.

[16] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," IEEE J. Biomed. Health Inform., vol. 22, no. 1, pp. 173–183, Jan. 2018.