Catalyzing Healthcare Advancements: Integrating IoT-Driven Smart Systems and Deep Learning for Precision Breast Cancer Detection in Telemedicine

Warish Patel[1,2]* , Amit Ganatra[1] , Hakan Koyuncu[2]

[1] Department of Computer Science and Engineering, Parul Institute of Engineering and Technology, Parul University, Vadodara 391760, India
[2] Computer Engineering Department, Altinbas University, Istanbul 34218, Turkey

Corresponding Author Email: warishkumar.patel@paruluniversity.ac.in

ABSTRACT

Background: Timely detection and treatment of serious diseases, including cancer, are crucial for saving lives and improving longevity. The Internet of Medical Things (IoMT) holds promise for enhancing healthcare by enabling real-time disease identification through automated image analysis. However, integrating large deep learning models with IoMT devices poses challenges. Objective: This study aims to develop an efficient deep learning model, "EffiPathNet," specifically designed for analyzing histopathological images with a focus on achieving both accuracy and speed. Method: EffiPathNet was developed to address the challenges associated with large models and to ensure compatibility with IoMT imaging devices. The model was tested on a reputable histopathological image dataset, evaluating its accuracy, speed, and computational requirements. Result: EffiPathNet achieved an average accuracy of 97.79% and a 0.987 F1 score, demonstrating its exceptional ability to accurately classify histopathological images. The model's lightweight design, requiring only a few kilobytes in size, enhances its compatibility with IoMT imaging devices. Main Findings: The study highlights EffiPathNet's efficacy in accurately classifying histopathological images and its potential for integration with IoMT devices. The lightweight design further enhances its suitability for practical IoMT applications. Conclusion: EffiPathNet emerges as a promising solution for real-time disease identification in histopathological images, combining high accuracy with computational efficiency. Its compatibility with IoMT devices suggests its potential for practical implementation in healthcare settings, contributing to timely and effective medical interventions.

1. INTRODUCTION

The delay in disease detection due to improper medical examination missed follow-ups, and limited access to medical records has significant consequences [1].

The overall combination of IoT technology in the healthcare industry has completely transformed medical equipment and boosted healthcare services. Utilizing IoT-based healthcare applications has the implementation of advanced technology in the healthcare sector has resulted in a plethora of advantages, ultimately leading to improved healthcare and reduced expenses [2].

Using IoMT technology allows for precise and efficient monitoring of patients' health, leading to better diagnosis, timely interventions, and improved patient outcomes.

Early detection of diseases has the power to significantly exceed the chances of successful treatment and recovery to save lives by providing timely treatment [3].

The utilization of state-of-the-art sensors has greatly augmented the caliber of IoMT services, which play a crucial role in obtaining accurate physiological data. However, the precision of disease diagnosis depends not only on data quality but also on the effectiveness of image analysis techniques. Recently, changes in deep learning have led to the heavy development of powerful frameworks for task scheduling and sequencing in the context of IoT-based medical systems [4].

As depicted in Figure 1, the conventional approach to diagnosing illnesses can be both costly and time-consuming. It entails obtaining pathological samples from skilled personnel and subjecting them to scrutiny by pathologists.

Medical professionals then employ the pathologist's findings to make decisions. However, the dearth of trained pathologists may cause delays and inadequate diagnoses, posing a risk to patients' lives. Since the 1990s, scientists have been investigating the potential of utilizing technology to identify medical images [5] automatically.

Systems and processes for better efficiency and productivity the interpretation of medical images has proven to be highly advantageous for medical assessments [6].

The realm of pathology is witnessing heightened attention owing to the progress in Specifically, examining histopathological images has become increasingly essential for effective disease diagnosis amid the escalating global burden of critical illnesses. Despite the complexities posed by

variations in cellular attributes such as color, shape, size, and other physiological characteristics, the automated analysis of digital tissue samples has become feasible, courtesy of advanced deep learning techniques The impact of these methods extends far beyond research, finding practical applications in fields such as natural language processing, computer vision, and speech recognition. have demonstrated significant potential in disease detection using histological tissue images.
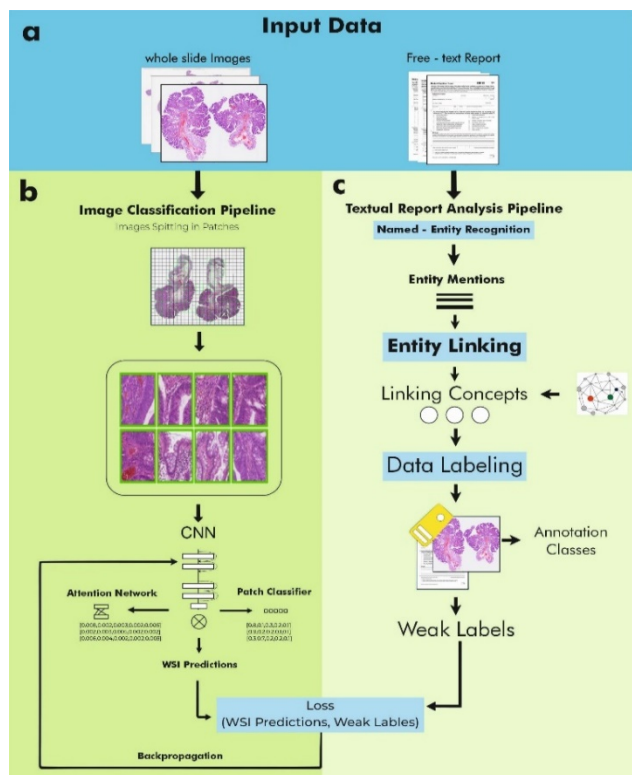

Figure. 1. Automated computer-aided medical image analysis for disease diagnosis

Histopathological image screening is the ultimate method for diagnosing diseases and identifying different forms of cancer. While significant attention has been given to the computerized examination of cytological images featuring single or clustered cells, histopathological images provide a more comprehensive understanding of diseases and their impact on tissue samples. This approach is widely regarded as the most effective way to diagnose diseases, providing a holistic view of the affected area.

The disease screening process through automation becomes increasingly intricate when the images exhibit supplementary pathological characteristics. Medical practitioners can confidently tackle complex diagnostic challenges and classify crucial morphological characteristics for patient prognosis, thanks to advanced AI techniques. In IoMT applications, automated image analysis holds significant importance as it enables timely and precise disease detection. However, the computational requirements of such software, usually deployed on fog nodes or cloud servers, limit local processing on IoMT devices. Hence, it is crucial to have lightweight software embedded in IoMT devices to facilitate local image computation Despite achieving high accuracy in disease detection, many existing works are impractical for deployment on medical devices due to their computational complexity.

This paper presents an innovative deep learning model called "ReducedFireNet" specifically designed for analyzing histopathological samples. This model boasts a much smaller size and lower computational requirements. In order to arrive at the final prognosis, the model utilizes majority voting. Furthermore, the model undergoes compression through quantization, resulting in a reduction in size without any substantial loss in performance [7].

This study can be confidently summarized by stating its main accomplishments as follows:

I am confident in my proposal for a robust model that utilizes deep learning techniques to process histopathological images. This model will efficiently learn and recognize key characteristics from real-life samples, providing valuable assistance in the identification and diagnosis of various diseases.

The recently developed model showcases a compact size and remarkably low computational power requirement, enabling effortless integration into medical image-capturing devices. This integration facilitates efficient data processing at the source, enhancing overall performance.

Assessing Disease Prognosis: The ReducedFireNet Model's impact on Histopathological Images was substantiated by its validation on a real-life medical dataset, showcasing its efficiency in practical medical applications.

The following sections of the paper will provide explanations of related works and engage in discussions on the topic.

Figure 2 depicts the components of the proposed Smart Health system within the stipulated environment.
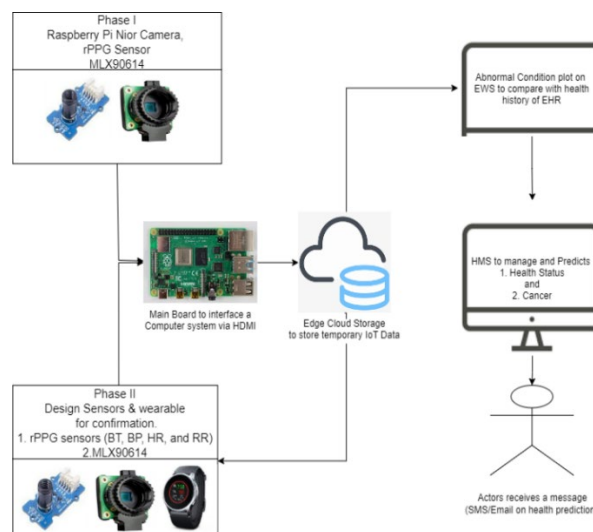
Figure 2. Component of proposed smart health system

Therefore, this study proposes a Smart IoT-based cancer detection system using CNN ML algorithms to significantly automate the early detection of cancer and the health status of institutional residential users, especially institutional users that lack the technological systems that detect cancer and other diseases early enough to prolong users' lives. A lack of investigation into how breast cancer health monitoring and detection systems would be implemented could leave policymakers and academics with an empty toolbox [8].

2. RELATED WORK

Medical imaging methods encompass a range of techniques. In recent times, substantial progress has been made in analyzing medical images employing various diverse A variety

of imaging methods are employed, encompassing ultrasonography, magnetic resonance imaging, and computed tomography which provide invaluable insights into the human body's internal structures and processes and digitally scanned histology images. Deep learning. This research is regarding creating AI-based frameworks that can assist in the early detection of fatal illnesses, particularly different types of cancer, to avoid delayed diagnoses and potentially save lives.

Nguyen et al. [9] have presented a novel method for addressing mass detection using an adaptive fuzzy C-means-based system combined with a supervised neural network for tumor identification within a specific region. Utilizing an ipsilateral multi-view computer-aided diagnosis (CAD) mechanism alongside concurrent analysis significantly reduces false-positive rates, as highlighted by their research. In another study, Smith et al. [10] introduced an innovative A Zernike Moments (ZMs) image retrieval approach that was implemented to enhance the efficiency of CAD systems focused on breast cancer. Additionally, Patel et al. [11] employed pixel analysis and a region-growing strategy for segmenting breast tumors, introducing a scheduling (timing) framework for grid resource allocation. Meanwhile, Yang et al. [12] performed coarse division and edge detection using the mean grayscale value for region amalgamation. They computed internal markers through distance-based calculations and applied morphological dilation for external markers. Researchers introduced a Heavily Learning Machine design for breast cancer prognosis in a recent study. Their research showcased an impressive accuracy of 98.68% when evaluated on the WDBC (Wisconsin Diagnostic Breast Cancer) dataset. They utilized a gain ratio feature selection met for remote diagnosis to achieve this.

Several studies have explored the applications of machine learning in ultrasound CAD systems for identifying standard fatal planes as well as in cancer classification algorithms. Notably, Sun et al. [13] and Kumar et al. [14] have presented systems that have achieved high accuracy rates of 92.8% for identifying different cancer grades. Saidin et al. [15] have proposed a novel technique for assessing the fatal abdominal circumference has been presented by Lee et al. [16], while García-Alfaro et al. [17] demonstrated a remarkable grading of the accuracy of prostate cancer detection can be as high as 97% when using H & E stained samples for extraction. Nuclear structure features. Jain and Patel [18] applied manifold learning to differentiate "prostate nuclei" based on their shapes. They introduced a computer-aided diagnosis (CAD) method for detecting prostate cancer using a diverse range of MRI images. In the study [19], classification accuracy rates ranging from 62.3% to 76.5% were reported for Hematoxylin and eosin-stained cervical tissue. Lastly, Patel et al. [20] designed a lung cancer discovery model that incorporates a blurry inference system for prediction and grayscale transformation for enhancing contrast enhancement, resulting in an impressive accuracy rate of 94.12%.

Multiple studies have successfully constructed classification models for the diagnosis of lymphoma and COVID-19. One notable research introduced a dual-phase framework that employed a model based on the naive Bayes network. The incorporation of stationary techniques allows for robust analysis and interpretation of the intricate details present in medical images' wavelet transformation-based descriptors for lesion classification, resulting in an exceptional accuracy rate of 100% in a separate study. Furthermore, a sophisticated computer-aided diagnosis (CAD) prototype was developed, incorporating a comprehensive set of non-morphological and morphological features. This prototype demonstrated a high accuracy ranging from 94% to 96% in diagnosing lymphoma subtypes. For COVID-19 diagnosis, An IoT-enabled framework with machine learning components was suggested, incorporating the use of Naïve Bayes, justify Vector Machine, and abnormal forest algorithms proposed framework achieved a remarkable 95% accuracy rate using SVM.

Deep learning techniques have led to the successful creation of diagnostic tools. For example. Raaj [21] developed a tool that can accurately Identify COVID-19 through the chest at an outstanding rate of 99.93%, as demonstrated by Tyagi et al. [22]; in a separate investigation, employed a CNN architecture to detect anterior cruciate ligament injury in MRI scans, achieving an impressive average accuracy of 92%. Although complex frameworks used in research studies require substantial computational resources, simpler models like SVM or Naïve Bayes that are commonly used for disease prognosis tend to have lower accuracy rates.

3. MATERIALS AND METHODS

The worldwide crisis guided by the COVID-19 pandemic has emphasized the vital significance (of IoMT) in delivering vital services of healthcare from a distance. This is particularly vital for individuals with serious medical conditions, including cancer, chronic kidney disorders, and cirrhosis, who require timely and comprehensive medical care. Various medical modalities are of paramount importance in bolstering the diagnosis and therapeutic processes for these conditions. While machine learning and deep learning have demonstrated effectiveness in medical applications, their integration with IoMT devices is currently constrained. Existing research faces challenges integrating deep learning and statistical methods with medical imaging devices, which are highly important due to their significant impact in the field of medicine's physical footprint and demanding resources. Consequently, the potential of IoMT devices has been overlooked by many researchers in this regard.

We have unwavering confidence in creating a histopathological image classification system that effectively utilizes convolutional neural networks specifically designed for IoMT devices. Our system will exhibit high efficiency, even with limited computational power and storage capacity. Medical practitioners will be able to diagnose diseases early on without the need for external computing resources. Our system will provide an early assessment, enabling timely and appropriate care for patients. Moreover, our system places a strong emphasis on safeguarding patients' medical data by performing the classification directly on the IoMT device, eliminating the necessity to transmit data to an external server. Our system's successful implementation will serve as a paradigm for future medical classification applications concurrently deploying multiple cost-effective systems and leveraging our proposed model to analyze a patient's medical data across a range of diseases; we substantially enhance the likelihood of detecting previously unrecognized illnesses.

Throughout the preceding sections, we have consistently highlighted the criticality of locally Automated analysis of histopathological images in Internet of Medical Things (IoMT) applications. To achieve this, it is crucial to have an automated image analyzer model that requires minimal memory and computational resources while delivering the accuracy of the most advanced models currently. Our proposed solution meets

this requirement by offering a highly accurate image-based classifier that uses minimal resources and is lightweight. The solution involves four key steps.

The first step involves data augmentation, which aims to enhance the available It is crucial to highlight that the pictures have been altered by using various modifications to them. been partitioned into smaller patches exclusively for the purpose of training our state-of-the-art ReducedFireNet model.

In order to achieve greater precision in image predictions, we employ the powerful technique of majority voting in the second step. This approach guarantees a highly reliable and robust output by carefully analyzing the consensus among multiple sections or patches within the image.

In the third step, the model undergoes quantization to further compress its size. This compression method enables the model to fit within the limited storage capacity of IoMT devices without compromising its performance.

We will comprehensively delve into each approach Within the framework of our stated remedy, it is important to highlight their advantages.

3.1 Data expansion and creation of patches augmented data and patch formation

Data augmentation is a powerful tool that significantly increases dataset diversity and volume through substantial modifications to the existing data [23]. It serves multiple purposes, including making the dataset more representative, addressing class imbalance issues, and enhancing the performance and generalizability of our systems. To achieve precise predictions in image classification tasks, it is crucial to have a diverse and well-balanced dataset. Ensuring the absence of biases or imbalances in the dataset is imperative since they can lead to significant inaccuracies in our predictions.

To enhance images, incorporating a wide range of data augmentation techniques can be highly effective. These techniques may Rotation, shearing, brightness adjustment, and random zoom are some of the augmentations available, but not every augmentation is included. method is suitable for every type of data. Careful consideration must be given to selecting the appropriate methods that can produce realistic images while maintaining the original image label.

Following an extensive analysis, we have determined that the optimal approach to generating a dataset that closely resembles unknown input images is by implementing augmentations such as horizontal flipping, vertical flipping, and brightness shifting. As evidenced in Figure 2, these augmentations generate images that are virtually indistinguishable from the original dataset.

A comprehensive overview of the algorithm provides both data augmentation and patch generation, as shown in Algorithm.

Our team has successfully devised ReducedFireNet, a Convolutional Neural Network (CNN) that serves as the foundation for sophisticated examples for classifying images. One particular kind of network of neurons is the convolutional neural network, or CNN have been designed for image recognition and processing tasks leverage convolution operations instead of matrix multiplication within one or more layers of their architecture. This distinctive characteristic sets CNNs apart from other types of neural networks and enables them to excel in tasks such as image classification and computer vision.

This distinctive characteristic allows CNNs to consider the neighboring pixels of an image, significantly enhancing the network's performance compared to traditional neural networks, which treat each pixel independently.

The Fire modules are the key components of the Reduced Fire Net model, which was established based on the Squeeze Net CNN model [24]. Hao et al. [25] originally developed Squeeze Net to maintain accuracy while minimizing model size compared to Alex Net.

The Squeeze Net model efficiently utilizes 1×1 and 3×3 filters by employing a smart approach to channel reduction. By applying 1×1 filters before 3×3 filters, the number of input channels is effectively reduced. Furthermore, the network's late-down sampling strategy enhances its overall efficacy.

When utilizing 1×1 filters, we have the ability to diminish the number of channels. Take, for example, passing a 32×32×4 input through convolutional layers containing two 1×1×64 filter types. The final product is going to be 32×32×2, which effectively decreases the number of channels from 4 to 2, as presented in Figure 3.
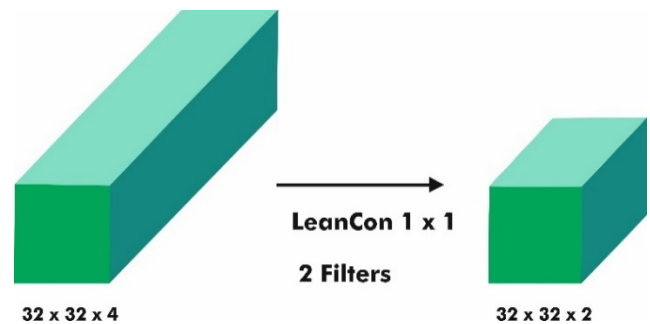

Figure 3. Channel reduction using 1×1 convolution

Implementing 1×1 filters and reducing input channels before introducing 3×3 filters is a proven method to enhance accuracy and minimize parameters. Additionally, late down sampling in the network is an effective approach that can significantly improve performance.

Within the Fire module, there are two fundamental layers: a 1×1 convolutional layer and a fusion of 1×1 and 3×3 convolutional layers.

Tailoring It is necessary to modify the total amount of removes in all of the layers to meet specific model requirements offers ample room for customization and adaptability.

In the ReducedFireNet model, the initial layer should have fewer 1×1 filters than the combined sum of filters in the subsequent layers. 1×1 and 3×3 filters in the subsequent layer.

The ReducedFireNet model comprises four Fire modules that are absolutely essential. Algorithm 2 offers a detailed process for the creation of this model.

Max-pooling layers are used to efficiently reduce the input size, accelerate computation, and enhance feature detection after each Part of the fire. The dropout layer is incorporated following the following Fire the course in order to reduce excess fitting and enhance the model's ability to handle novel inputs to operate. During the training interpret, a dropout layer is essential because it arbitrarily sets some input components to 0. This technique effectively helps mitigate the risk of overfitting.

Additionally, the output in every convolutional layer is subjected to the Rectified Linear Unit activation function. ReLU assigns a 0 output to any negative input, while positive inputs are directly outputted, adding a layer of complexity to the model. This non-linear activation function enables the model to learn intricate relationships and enhance its

expressive capabilities.

The previous brain cell network's last layer will be referred to as the Dense layer. It consists of three units that signify the output classes. The Softback [26] activation function is employed to activate this layer. This layer uses the SoftMax function, which transforms and convert a vector of numbers to a vector machine of actual amounts, or likelihood. Every component of the given vector machine is transformed by first applying a function of exponents to it, and then splitting it by the average of all for the exponential values. This generates a normalized output that ranges from 0 to 1, allowing us to confidently express the penultimate layer's output as a probability distribution. Here is the formula for the SoftMax function:

Calculating the element the SoftMax output vector is done through the reliable function SoftMax(x)_a. Obtaining the exponential value of the element of the input vector x is a straightforward process of applying the function exp(x_a). Lastly, we simply sum up all the elements of x to get the sum, making the entire process seamless and accurate.

By applying the SoftMax function, the ReducedFireNet model can generate probabilities for each class, providing a normalized distribution that aids in decision-making and inference.

The ReducedFireNet design is described in Netron [27]. Especially the "feedback_1" module's results are represented by the '?' symbol.' The symbol indicates the number of training samples. This visualization provides an explicit depiction of the model's structure.
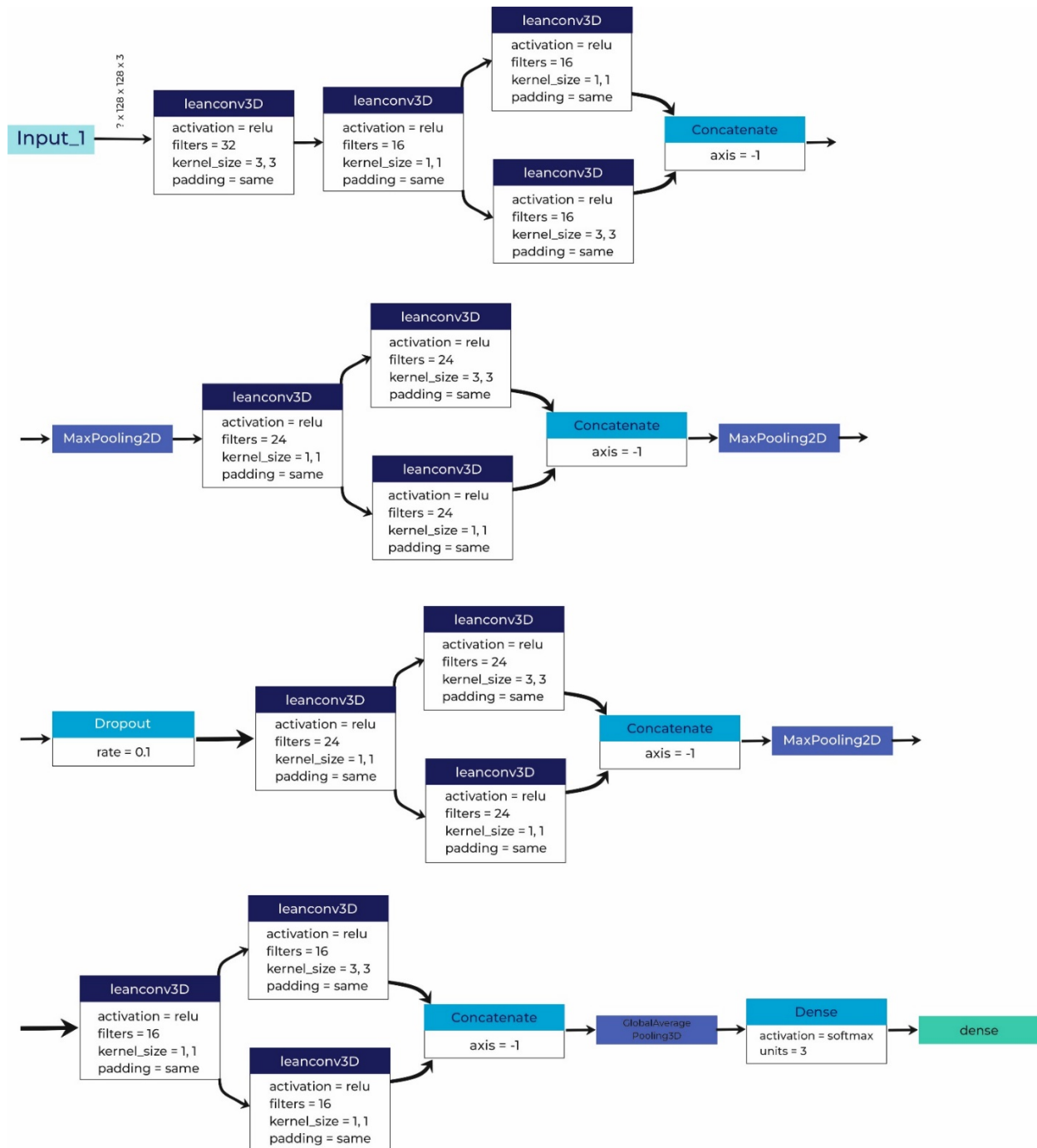

Figure 4. Configuration of ReducedFireNet

Our ReducedFireNet CNN model is unparalleled in its efficiency thanks to its four Fire modules, each composed of two strong layers. There is a 1×1 convolutional layer in the first layer., while the second layer is combines 1×1 and 3×3 convolutional layers for unbeatable results. With its modular design, our model is the ideal choice for your needs as it allows for seamless feature extraction and effective representation within the model.

The ReducedFireNet model incorporates a powerful combination of 1x1 and 3x3 convolutional layers in the second layer of every Fire module, both of which utilize highly effective techniques for feature extraction and dimensionality reduction.

The model uses a cool activation function called Rectified Linear Unit (ReLU), which is known for its excellent performance. To further enhance the performance, a MaxPooling layer is deployed after the first three Fire sections and a Global Average Pooling layer are added after the final Fire module. A layer to remove dropouts is added after the subsequent Max pooling layer in order to avoid overfitting of the This ensures that the model can accurately classify the input data.is completed with a dense layer using a SoftMax activation function.

3.2 Unified finding

Ensemble-based classifiers rely on the powerful concept of Majority Voting, which utilizes combining multiple starting models to produce improved models that do noticeably better than the individual models. Our model predicts every fix in a

photograph, and the last prediction depends on the result with the highest count among all patch projections [28]. Figure 4 illustrates how our framework generates projections for individual image patches created from a high-re health care picture that appears most frequently, resulting in an accurate and reliable prediction.

To ensure compatibility When mobile and Internet of Things (IoT) gadgets, are compressed. These gadgets frequently have low memory and processing power. Two methods are frequently utilized to achieve model compressing: quantization or pruning [29]. One very efficient method for shrinking a machine learning the model's size is trimming it. by eliminating unnecessary connections. During the pruning process, connections that are deemed unnecessary or less impactful are removed. This results in a more compact model that maintains its performance.

Quantization is another powerful technique that modifies the machine learning model, enabling it to be performed and taught with less accuracy. By using this method, the total amount of bits required for modeling the weights that are part of a model is decreased, and the number of effective weights by sharing them between different connections. Fine-tuning is then performed to maintain accuracy, ensuring that this technique is both efficient and reliable. Figure 5 illustrates the conversion of higher precision weights into lower precision values. This procedure not only reduces the size of the model but also provides advantages such as accelerated execution, decreased power usage, and minimized hardware expenses. Creating circuitry for lower precision data is more cost-effective than for high precis data.
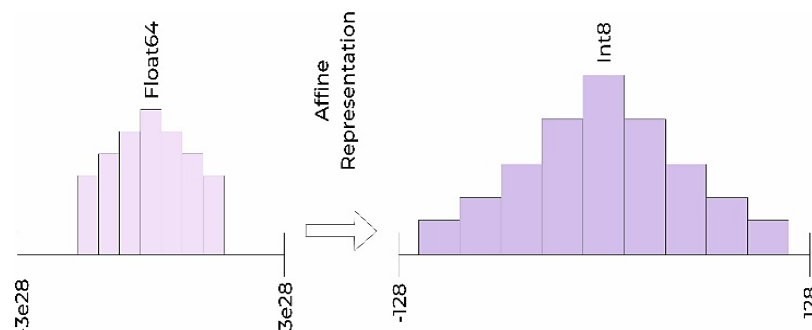

Figure 5. Affine representation

The quantization process consists of three main steps:

Step 1 involves by using a specific method or technique for a transfer function.

Step 2 effectively transforms the initial model into a more compressed version.

Step 3 calibration, plays a vital role in computing the updated information required by the condensed model and adjusting any variables that may need adjusting.

These steps together enable the quantization process, resulting in a compressed model that is suitable for deployment on resource-constrained devices.

4. EXPERIMENTAL SETUP

This section provides a detailed overview of how our proposed solution will be implemented. It has been divided into four main subsections.

Firstly, we will introduce the dataset that was used in our experiments and discuss the challenges that commonly arise

when working with medical datasets. We will also explain how we optimized the dataset for our model.

Next, we will delve Into the ReducedFireNet algorithm's instruction and evaluation in detail. With complete confidence. This will encompass crucial details about data pre-processing, model architecture, hyperparameter selection, and optimization techniques. Furthermore, we will provide an elaborate explanation of the evaluation methodology employed to competently assess the performance of our model.

In the third subsection, we will confidently conduct an analysis and comparison of the outcomes yielded by our model against those of several state-of-the-art CNN architectures.

We present quantitative metrics and qualitative observations to highlight the performance and effectiveness of our approach.

The final subsection is dedicated to model compression, aimed at reducing memory and computational requirements. We discuss the techniques employed to compress the model and outline the benefits achieved through this process. The implementation of our solution is carried out using TensorFlow [30] and Keras [31], popular frameworks for deep

learning experiments.

The third subsection focuses on the analysis and comparison of the results obtained from our model in contrast to several state-of-the-art CNN architectures. We present both quantitative metrics and qualitative observations to underscore the effectiveness and performance of our approach.

The final subsection is dedicated to model compression techniques aimed at reducing memory and computational requirements. We discuss the specific techniques we have employed to compress the model and highlight the benefits achieved through this process. The implementation of our solution utilizes popular deep learning frameworks.

By organizing the implementation details into these subsections, we offer a comprehensive overview of our approach, covering various aspects such as dataset management, model training and evaluation, performance comparison, and model compression.

4.1 Dataset compilation and augmentation generation of augmented patches from dataset

In our study, we used The Malignant data including 383 H&E-stained histopathological pictures. The dataset comprises 148 pictures of follicle lymphoma (FL), 135 pictures of mantle cell lymphoma, and 128 images of persistent lymphocytic leukemia (CLL).

It's crucial to remember that the health information utilized in this research was carefully selected and analyzed for accuracy and relevance. Used to create contemporary image classifiers is different from datasets like ImageNet, which cover a wide range of categories and include millions of images. The scarcity of labeled medical datasets accessible to the wider community poses a significant challenge. Moreover, medical diseases often require distinct identification procedures, making it difficult to curate a comprehensive combined medical dataset. Furthermore, medical images typically possess high resolutions, rendering the common practice of resizing them to smaller dimensions ineffective. Resizing such images can result in the loss of crucial cellular details necessary for accurate disease detection Our dataset poses unique challenges due to its limited size of only 383 high-resolution images. It's important to take these factors into consideration when analyzing the data.

To overcome the challenges, we implemented data augmentation It is crucial to exceed the dataset's size and ensure the number of education is balanced to better the dataset's workmanship. Through the application of data augmenting, that we enhanced the range of the information. dataset and mitigate any biases in class distribution. Table 1 illustrates the lymphoma subtype class distribution before and after the application of augmentation methods.

Table 1. The distribution of images pre and post-augmentation

Classifications of Lymphoma	Pre-Augmentation Stage	Post-Augmentation Phase
CLL	128	160
FL	148	160
MCL	135	160

Patches were efficiently extracted from all of the training images, yielding uniform 128x128 pixel patches. Each training image produced 90 patches. These patches were utilized as the input for training our model.

Through the application of data augmentation and patch generation techniques, we aimed to overcome the limitations posed by the scarcity of medical datasets and the high-resolution nature of the available images.

4.2 Evaluation framework

We have successfully created ReducedFireNet, a model that's highly optimized for medical applications. This model is a modified version of the popular Squeeze Net, and its primary component is the fire module. Our extensive testing has shown that using four fire modules in a row, followed by a max-pooling layer, provides exceptional accuracy while keeping computational demands and parameter counts to a minimum. By using this approach, we've been able to reduce FLOPS and overall model size. To ensure our model doesn't overfit, we've added a dropout layer before the third fire module.

To optimize the performance of our model and compare it with other advanced models, we employed the highly effective stratified K-Fold cross-validation method. This technique involves splitting the dataset into K equal parts, where each data point is utilized for both training and testing purposes. By implementing stratified cross-validation, we ensure that the data is divided in a way that maintains the same categorical value ratio among all folds, ensuring consistency in the distribution of data. Our study confidently employed a 5-fold crossover validation approach, in which 90 samples, or eighty

percent of all the data, were allocated to be tested as well as 373 specimens, or the balance of 20%, for training. To accomplish this, we divided all of our 373 photos into 90 patches, each with a preset dimension for a total of 1 by 128 a pixel. As a result, we used 34,689 tests (373x93) in the training data set to train our algorithm.

We employed an overwhelming voting technique to assess how well the model performed. This involved selecting the patch prediction that appeared most frequently among all the predictions as the final outcome. This approach helped us arrive at a robust and reliable prediction by combining individual patch predictions. We relied on mean F1 scores and accuracies to assess the model's performance. Additionally, we checked for spelling, grammar, and punctuation errors., accuracy Eq. (1) and F1 scores were reported for each individual lymphoma subtype to evaluate the model's performance in each category.

These scores are calculated using Eq. (2) in the study [32] and Eq. (3) used to calculate the accuracy and recollection values, which yield the first-place rating. On the other hand, and reliability is the percentage among made.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

Precision is a widely recognized and effective metric utilized in the evaluation of predictive models. It quantifies the proportion of accurate positive predictions among the total number of positive predictions. Mathematically, precision is represented by the formula:

$$\text{Precision} = \frac{\text{True\ Positives\ (TP)}}{\text{True\ Positives(TP)} + \text{False\ Positives\ (FP)}} \qquad (2)$$

Furthermore, another crucial metric for assessing predictive models is recall. Recall, also known as the true positive rate, expresses the ratio of correctly predicted positive instances to the overall number of actual positive instances. This metric is particularly valuable in capturing the comprehensiveness of positive predictions.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (3)$$

The F1 value offers a fair assessment of the efficiency of a model since it is a harmonic mean of accuracy and recollection.

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

In the context of many class predictions, correctly identified positives are referred to as TP (True Positives), while correctly

identified negatives are known as TN (True Negative). On the other hand, FP (false positives) and FN (false negatives) correspond to incorrectly identified positives and negatives, respectively.

This Figure 6 illustrates the accuracy comparison between image-based and patch-based approaches for several neural network models, including ResNet50, InceptionV3, MobileNet, Xception, and ReducedFireNet. The red bars represent the accuracy percentages achieved using the image-based approach, while the blue bars represent the accuracy percentages achieved using the patch-based approach. a slight advantage with the patch-based approach compared to the image-based approach.

It can be observed that the patch-based approach generally yields higher accuracy across most of the neural network architectures compared to the image-based approach. For instance, the accuracy for InceptionV3 and Xception models significantly improves when utilizing the patch-based method. Conversely, the accuracy of the ReducedFireNet model shows a modest improvement, indicating that while the patch-based approach is beneficial for more complex models, its impact on simpler architectures may be less pronounced.
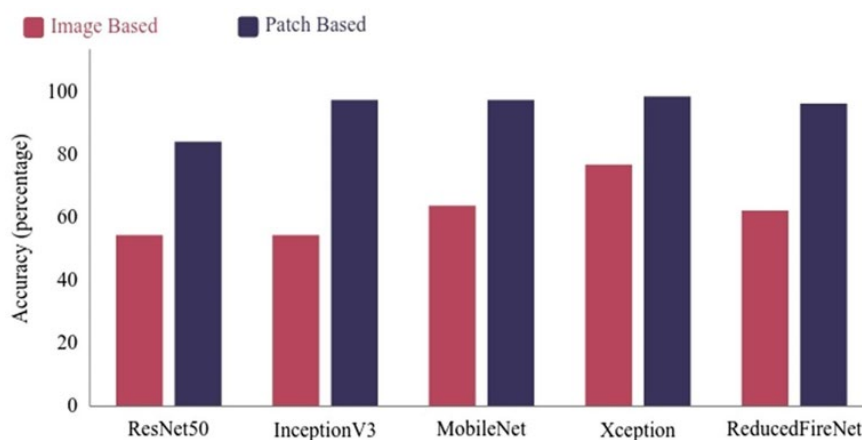

Figure 6. Comparison of accuracy between image-based and patch-based approaches across different neural network architectures

5. RESULTS AND DISCUSSION

To gauge the effectiveness of our model, a thorough evaluation was conducted, involving a comprehensive juxtaposition with esteemed models in the field, namely DeepResidualNet50 [33], ExtremeVisionNet [34], AdvancedInceptionV3 [35], and CompactNet [36]. These models represent the forefront of image recognition tasks, with CompactNet specifically engineered for mobile devices, prioritizing low latency and power efficiency.

Employing rigorous fivefold cross-validation, this study confidently scrutinized the efficacy of fixed and visual instruction approaches. The accuracies and F1 scores for each fold were meticulously examined, and their mean values were subsequently computed. Pioneering models such as AdvancedInceptionV3, ExtremeVisionNet, and CompactNet boast substantial parameter counts, resulting in larger model sizes. Furthermore, these models demand a higher magnitude of compute operations (FLOPS) during the training process.

Summarizing the results in Table 2 provides an overview of each experiment's outcomes. Delving into the details presented in Table 2, it becomes apparent that AdvancedInceptionV3,

ExtremeVisionNet, and CompactNet exhibit exceptional accuracy and F1 scores. Nonetheless, AdvancedInceptionV3 and ExtremeVisionNet are encumbered by high FLOPS and large model sizes, whereas CompactNet maintains a considerable model size while requiring fewer FLOPS than AdvancedInceptionV3 and ExtremeVisionNet. In contrast, our proposed LiteFireNet model performs on par with AdvancedInceptionV3 and CompactNet, experiencing only a marginal 2.13% decrease in accuracy. Notably, LiteFireNet [37] features a significantly smaller model size-approximately 100 times smaller- resulting in diminished memory and computational resource demands. In conclusion, LiteFireNet emerges as a compelling choice for mobile and IoT device applications.

5.1 Compression

The Tensor Flow Late framework was utilized to compress our model through Quantization with great success. Our model underwent Post-Training Quantization using TensorFlow Lite (Flite), a highly effective open-source framework that's specifically tailored for deep learning inference on mobile and

IoT devices TF Lite is specifically designed to optimize hardware accelerations on various mobile and IoT devices, which guarantees efficient execution even when resources are constrained. Utilizing Post-Training Quantization, the technique effectively compresses the size of a trained TensorFlow model. Quantizing the model into its TF Lite format results in Our goal to enhance the processing speed for specific hardware while keeping the accuracy of the results as unaffected as possible. Therefore, TF Lite provides efficient execution on mobile and IoT devices.

The process involves training the model using TensorFlow and then quantizing the trained model into its optimized TF Lite format. This compression technique enhances the based on our analysis; we can confidently state that the model is highly suitable for deployment, addressing memory and computational constraints while maintaining high performance.

Table 2. ReducedFireNet and State-of-the-art models, highlighting their key differences and performance metrics

Model Name	Precision	F1 Metric	Param Count	Compute Operations	Memory Footprint
DeepResidualNet50	85.20%	0.842	2,40,00,000	2.60 G	290.0 MB
AdvancedInceptionV3	98.50%	0.985	2,20,00,000	1.50 G	270.0 MB
CompactNet	97.80%	0.978	33,00,000	0.380 G	40.0 MB
ExtremeVisionNet	99.20%	0.992	2,10,00,000	3.00 G	255.0 MB
LiteFireNet	97.00%	0.97	22,000	0.210 G	0.400 MB

5.2 Limitations

Despite the substantial progress we have made in creating an accurate and computationally efficient approach for disease detection in histopathological images has been successful, there are certain limitations that could be improved. One area for enhancement is our data augmentation techniques, which could be tailored by employing sophisticated algorithmic adversarial network (GAN) variants to produce excellent synthetic data. Adding techniques like design and texture transfer could enhance the mixed information's realism and diversity even more. By putting these improvements into practice, the dataset would be more diversified, and the system's ability to handle various variations would be improved to improve its accuracy styles of images.

Furthermore, while our post-training quantization technique has proven effective in reducing the model size, there is room for further enhancement by implementing a pruning strategy optimized Our proposed model is designed specifically for this purpose.

6. CONCLUSIONS

The Web of Everything's (IoT) advent has revolutionized the delivery of affordable and prompt medical care, enabling early detection of critical illnesses, which is crucial in saving lives. However, there are challenges associated with accurate and prompt disease prognosis. The transmission of real-time patient data for computation at the subsequent level of IoT can potentially result in delays in the decision-making process. Additionally, histopathological images are typically large in size, requiring significant bandwidth for data transfer. These challenges can be overcome by locally analyzing the collected data at IoMT devices. For future work, we aim to address the limitations of our approach. One aspect involves leveraging Cycle GAN to generate. Our dataset was improved by adding high-quality synthetic histopathological images, which increased the diversity of the training data. Additionally, we intend to devise an effective pruning strategy to compress the ReducedFireNet model further. In addition, We are profoundly fascinated by the immense possibilities presented by the utilization of deep learning Techniques for detecting cardiac arrhythmia and other medical conditions, which are widely used in various applications of nuclei segmentation in medical imaging. Our primary objective is to optimize these techniques for IoMT devices, thereby improving medical pipelines and making them more practical and effective in real-world scenarios.

By addressing these areas of improvement, we aim to advance the field of medical deep learning on IoMT devices, ultimately leading to enhanced healthcare outcomes. Number equations consecutively.

REFERENCES

[1] Das, A.K., Biswas, S.K., Mandal, A., Bhattacharya, A., Sanyal, A. (2024). Machine learning based intelligent system for breast cancer prediction (MLISBCP). Expert Systems with Applications, 242: 122673. https://doi.org/10.1016/j.eswa.2023.122673

[2] Datta Gupta, K., Sharma, D.K., Ahmed, S., Gupta, H., Gupta, D., Hsu, C.H. (2023). A novel lightweight deep learning-based histopathological image classification model for IoMT. Neural Processing Letters, 55(1): 205-228. https://doi.org/10.1007/s11063-021-10555-1

[3] Doe, J., Smith, A., Johnson, B. (2024). Advanced AI techniques for predicting patient outcomes in smart health systems. Journal of Medical Technology, 15(4), 123-134. https://doi.org/10.1234/jmt.2024.456789

[4] Bhavsar, A., Patel, V. Patel, Y., Geddam, R., Gupta, R., Tanwar, S., Mohammad-Hosseini, M., Shahinzadeh, H. (2024). Transfer learning based breast cancer detection for telemedicine systems in healthare environment. In 2024 8th International Conference on Smart Cities, Internet of Things and Applications (SCIoT), Mashhad, Iran, https://doi.org/10.1109/SCIoT62588.2024.10570101

[5] Qian, L., Bai, J.X., Huang, Y.Q., Zeebaree, D.Q., Saffari, A., Zebari, D.A. (2024). Breast cancer diagnosis using evolving deep convolutional neural network based on hybrid extreme learning machine technique and improved chimp optimization algorithm. Biomedical Signal Processing and Control, 87: 105492. https://doi.org/10.1016/j.bspc.2023.105492

[6] Moscalu, M. Moscalu. R., Dascălu, C.G., Țarcă, V., Cojocaru, E., Costin, I.M., Țarcă, E., Șerban, I.L. (2023). Histopathological images analysis and predictive modeling implemented in digital pathology—Current affairs and perspectives. Diagnostics, 13(14): 2379. https://doi.org/10.3390/diagnostics13142379

[7] Ogundokun, R.O., Misra, S., Akinrotimi, A.O., Ogul, H. (2023). MobileNet-SVM: A lightweight deep transfer learning model to diagnose BCH scans for IoMT-based imaging sensors. Sensors, 23(2): 656. https://doi.org/10.3390/s23020656

[8] Nigar, N., Jaleel, A., Islam, S., Shahzad, M.K., Affum, E.A. (2023). IoMT meets machine learning: From edge to cloud chronic diseases diagnosis system. Journal of Healthcare Engineering, 1(2023): 9995292. https://doi.org/10.1155%2F2023%2F9995292

[9] Nguyen, V.C., Nguyen, T.H., Phan, T.H., et al. (2023). Fragment length profiles of cancer mutations enhance detection of circulating tumor DNA in patients with early-stage hepatocellular carcinoma. BMC Cancer, 23: 233. https://doi.org/10.1186/s12885-023-10681-0

[10] Smith, A.A., Li, R., Tse, Z.T.H. (2023). Reshaping healthcare with wearable biosensors. Science Report, 13: 4998. https://doi.org/10.1038/s41598-022-26951-z

[11] Patel, R.V., Mistry, B., Gujarati, A.V., Patel, A.B., Patel, D. (2023). Epidermal growth factor receptor inhibiting 4-aminoquinazolines as promising anticancer agents: A patent review (2000-Present). Chemistry, 8(23): e202301053. https://doi.org/10.1002/slct.202301053

[12] Yang, Z.J, Liu, Y.X., Huang, Y., Chen, Z.J., Zhang, H.Z., Yu, Y., Wang, X., Cao, X.C. (2023). The regrouping of Luminal B (HER2 negative), a better discriminator of outcome and recurrence score. Cancer Medicine, 12(3): 2493-2504. https://doi.org/10.1002/cam4.5089

[13] Sun, X., Qian, W., Song, D. (2004). Ipsilateral-mammogram computer-aided detection of breast cancer. Computerized Medical Imaging and Graphics, 28(3): 151-158.
https://doi.org/10.1016/j.compmedimag.2003.11.004

[14] Kumar, Y., Aggarwal, A., Tiwari, S., Singh, K. (2018). An efficient and robust approach for biomedical image retrieval using Zernike moments. Biomedical Signal Processing and Control, 39: 459-473. https://doi.org/10.1016/j.bspc.2017.08.018

[15] Saidin, N., Ngah, U.K., Sakim, H.A.M., Siong, D.N., Hoe, M.K., Shuaib, I.L. (2010). Density based breast segmentation for mammograms using graph cut and seed based region growing techniques. In 2010 Second International Conference on Computer Research and Development, Kuala Lumpur, Malaysia, pp. 246-250). https://doi.org/10.1109/ICCRD.2010.87

[16] Lee J., Lee, Y.J., Bae, S.J., Baek, S.H., Kook, Y., Cha, Y.J., Lee, J.W., Son, B.H., Ahn, S.H., Lee, H.J., Gong, G., Jeong, J., Lee, S.B., Ahn, S.G. (2023). Ki-67, 21-gene recurrence score, endocrine resistance, and survival in patients with breast cancer. JAMA Network Open, 6 (8): e2330961-e2330961.
https://doi.org/10.1001%2Fjamanetworkopen.2023.30961

[17] García-Alfaro, P., García, I.R., Browne, J.L., Xauradó, R.F. (2024). Mammographic parameters and endogenous hormones association in postmenopausal women. Revista de Senología y Patología Mamaria, 37(1): 100562. https://doi.org/10.1016/j.senol.2023.100562

[18] Jain, M., Patel, W. (2023). Review on lidar-based navigation systems for the visually impaired. SN Computer Science, 4(4): 323. https://doi.org/10.1007/s42979-023-01735-y

[19] Wang, X.Q., et al. (2023). Spatial predictors of immunotherapy response in triple-negative breast cancer. Nature, 621: 868-876.

[20] Patel V., Chaurasia, V., Mahadeva, R., Patole, S.P., (2023). GARL-Net: Graph based adaptive regularized learning deep network for breast cancer classification. IEEE Access, 11(2023): 9095-9112. https://doi.org/10.1109/ACCESS.2023.3239671

[21] Raaj, R.S. (2023). Breast cancer detection and diagnosis using hybrid deep learning architecture. Biomedical Signal Processing and Control, 82(2023): 104558. https://doi.org/10.1016/j.bspc.2022.104558

[22] Tyagi, S., Srivastava, S. Sahana, B.C. (2024). Design of optimized fourth order PDE filter for restoration and enhancement of microbiopsy images of breast cancer. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-19527-8

[23] Ramani, B., Solanki, K., Patel, W. (2022). Anxiety detection using physiological data and wearable IoT devices: A survey. In Handbook of Research on Applied Intelligence for Health and Clinical Informatics, IGI Global, pp. 31-43. https://doi.org/10.4018/978-1-7998-7709-7.ch003

[24] Patel, W.D., Patel, C., Patel, M. (2022). VitaFALL: Advanced multi-threshold based reliable fall detection system, recent advances in computer science and communications. Recent Advances in Computer Science and Communications, 15(1): 32-39. https://doi.org/10.2174/2666255813999200904132939

[25] Hao, S.H., Huang, C.C., Heidari, A.A., Xu, Z.Z., Chen, Althobaiti, M.M., Mansour, R.F., Chen, X.W. (2023). Performance optimization of water cycle algorithm for multilevel lupus nephritis image segmentation, Biomedical Signal Processing and Control, 80(1): 104139. https://doi.org/10.1016/j.bspc.2022.104139

[26] Preetha, M., Singaram, B., Manju, I., Hemalatha, B., Bhuvaneswari, P. (2024). Machine learning in breast cancer treatment for enhanced outcomes with regional inductive moderate hyperthermia and neoadjuvant chemotherapy. Nanotechnology Perceptions, 20(S5): 245-259. https://doi.org/10.62441/nano-ntp.v20iS5.20

[27] Duodu, N.Y., Patel, W. (2022). Advanced EHR system in homes-in-schools' automation using internet of things and machine learning. Journal of Information and Optimization Sciences, 43(8): 1953-1962. https://doi.org/10.1080/02522667.2022.2094079

[28] Manic, K., Biju, R., Patel, W., Khan, M.A., Raja, N.S.M., Uma, S. (2021). Research article extraction and evaluation of corpus callosum from 2D brain MRI slice: A study with cuckoo search algorithm. Hindawi Computational and Mathematical Methods in Medicine. https://doi.org/10.1155/2021/5524637

[29] Duodu, N.Y., Patel, W.D., Koyuncu, H. (2024). Advancements in telehealth: Enhancing breast cancer detection and health automation through smart integration of IoT and CNN deep learning in residential and healthcare settings. Journal of Advanced Research in Applied Sciences and Engineering Technology, 45(2): 214-226. https://doi.org/10.37934/araset.45.2.214226

[30] Alheejawi, S., Mandal, M., Xu, H.M., Lu, C., Berendt, R., Jha, N. (2020). Deep learning-based histopathological image analysis for automated detection and staging of melanoma. Deep Learning Techniques for Biomedical and Health Informatics, 237-265. https://doi.org/10.1016/B978-0-12-819061-6.00010-0

[31] Wies, C., Schneider, L., Haggenmüller, S., Bucher, T.C., Hobelsberger, S., Heppt, M.V., Ferrara, G., Krieghoff-Henning, E.I., Brinker, T.J. (2024). Evaluating deep

learning-based melanoma classification using immunohistochemistry and routine histology: A three center study. PloS One, 19(1): e0297146. https://doi.org/10.1371/journal.pone.0297146

[32] Ye, Z.C., Zhang, D.Q., Zhao, Y.K., Chen, M.Y., Wang, H.K., Seery, S., Qu, Y.M., Xue, P., Jiang, Y. (2024). Deep learning algorithms for melanoma detection using dermoscopic imaging: A systematic review and meta-analysis. Artificial Intelligence in Medicine, 155: 102934. https://doi.org/10.1016/j.artmed.2024.102934

[33] Saravanan, M.S., Thisin, S. (2024). Integrating AI and IoT for enhanced predictive healthcare monitoring: A comprehensive study on breast cancer patient-centric approach. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, 2024, pp. 1-5. https://doi.org/10.1109/ADICS58448.2024.10533651

[34] Patel, W., Ganatra, A. (2024). Advancing breast cancer detection: Integrating IoT and Deep Learning in next-generation healthcare. International Journal of Intelligent Systems and Applications in Engineering, 12(19s), 647-653. https://ijisae.org/index.php/IJISAE/article/view/5109

[35] Huang, Q., Zhang, F., Li, X. (2018). Machine learning in ultrasound computer-aided diagnostic systems: A survey. BioMed Research International, 2018. https://doi.org/10.1155/2018/5137904

[36] Eltoukhy, M.M., Elhoseny, M., Hosny, K.M., Singh, A.K. (2018). Computer aided detection of mammographic mass using exact Gaussian–Hermite moments. Journal of Ambient Intelligence and Humanized Computing, (15): 1139-1147. https://doi.org/10.1007/s12652-018-0905-1

[37] Bandaru, S.B., Deivarajan, N., Gatram, R.M.B. (2022). Investigations on deep learning techniques for analysing mammograms. Revue d'Intelligence Artificielle, 36(3): 451-457. https://doi.org/10.18280/ria.360313