JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ
العدد الواحد والثلاثون
(559) (537)

تكامل الذكاء الآلي وإنترنت الأشياء في الرعاية الصحية الذكية للكشف عن أمراض الصوت (IoT)

م.م قيس نعمه ابراهيم جاسم
kaisn2136@gmail.com
م.م يوسف حمد عيفان عبد
yousifxyz@gmail.com
م.م احمد مطر عواد محمد
amaacs2@gmail.com
مديرية تربية الانبار

المستخلص:

يمثل توظيف تقنيات الذكاء الاصطناعي (AI) وإنترنت الأشياء (IoT) في الرعاية الصحية الذكية اتجاهاً واعداً للابتكار. لقد أسهمت التطورات الأخيرة في تقنيات التعلم العميق (Deep Learning) في دفع عجلة أتمتة أنظمة التشخيص والتصنيف بشكل ملحوظ. كما أن ظهور شبكات الجيل الخامس (G٥) في مجال الاتصالات اللاسلكية قد وفر سرعة أكبر وموثوقية أعلى في نقل البيانات، مما سرّع من تبني حلول الرعاية الصحية الذكية. وقد أبرزت جائحة كوفيد-١٩ الأهمية البالغة لمثل هذه الأنظمة. تؤثر أمراض الصوت على شريحة واسعة من الأفراد حول العالم، غير أن التشخيص المبكر لها يجعل علاجها أكثر فاعلية. في هذه الدراسة، نقترح إطاراً للرعاية الصحية الذكية مخصصاً للكشف عن أمراض الصوت. تُسجَّل إشارات الصوت والـ EGG باستخدام أجهزة إنترنت الأشياء، بما في ذلك الميكروفونات وأجهزة التخطيط الكهربائي للحنجرة (Electroglottography). ثم تُحوَّل هذه الإشارات إلى أطياف ميل (Mel-spectrograms) وتُحلَّل بواسطة شبكات عصبية تلافيفية (CNNs) مدرَّبة مسبقاً. بعد ذلك، تُدمَج الخصائص المستخرجة باستخدام شبكة LSTM ثنائية الاتجاه (BiLSTM) لالتقاط الارتباطات الزمنية. وقد جرى تقييم الإطار المقترح باستخدام قاعدة بيانات ساربروكن الصوتية. وتبين من النتائج التجريبية أن المدخلات ثنائية النمط تحقق أداءً أفضل من أحادية النمط، حيث بلغت دقة التصنيف ٩٥.٦٥%. تُظهر الدراسات أن المدخلات ثنائية النمط (Bimodal Inputs) تحقق أداءً أفضل من المدخلات

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

أحادية النمط، وقد حقق النهج المقترَح دقة (Precision) بلغت 95.65%. الكلمات المفتاحية: الشبكة العصبية التلافيفية، الرعاية الصحية الذكية، التعلم العميق، الذاكرة طويلة المدى قصيرة المدى (LSTM), الكشف عن أمراض الصوت .

Integration of Machine Intelligence and Internet of Things in Smart (Healthcare for Voice Pathology Detection (IoT

QAYS NEAMAH IBRAHIM
kaisn2136@gmail.com
Yousif Hamad Efan
yousifxyz@gmail.com
Ahmed Mutar Awad
amaacs2@gmail.com
Directorate of Education in Anbar, Iraq

Abstract:

The use of Artificial Intelligence (AI) together with the Internet of Things (IoT) in smart healthcare represents a promising direction for innovation. Recent progress in Deep Learning techniques has significantly advanced the automation of diagnostic and classification systems. Furthermore, the emergence of 5G wireless networks has provided faster and more reliable data transmission, accelerating the deployment of intelligent healthcare solutions.

The COVID-19 pandemic highlighted the critical importance of such systems. Voice disorders affect a large number of individuals worldwide, yet early detection can make them highly treatable. In this study, we introduce a smart healthcare framework for the detection of voice pathology. Voice and electroglottography (EGG) signals are recorded using IoT-enabled devices, including microphones and EGG sensors. These signals are converted into Mel-spectrograms and analyzed through pre-trained convolutional neural networks (CNNs). The extracted features are then integrated using a bidirectional LSTM (BiLSTM) network. The proposed framework is evaluated with the Saarbrücken speech database. Experimental results

indicate that bimodal inputs provide better performance than unimodal ones, achieving a classification accuracy of 95.65%.

Keywords: Convolutional neural network, smart healthcare, deep learning, LSTM (Long Short-Term Memory), voice pathology detection .

I.Introduction

Many people nowadays have voice disorders due to overusing their vocal cords. Teachers, students, artists, lawyers, and the like are among those who often encounter these issues (Muhammad et al., 2017, p. 69-73). People's voices contain an enormous amount of information, and as a consequence, they might express much too much about their well-being. Automated speech pathology diagnosis and speaker identification are only two examples of where this information might be used. As a result, many academics are drawn to speech pathology to learn more about the many abnormalities of the voice. Vocal folds with abnormal development of bulk or tissue produce a voice tone distinct from that of a typical person (Ali et al., 2016, p. 757). Vocal disorders may result from an abnormal expansion of the vocal folds, vocal fold cyst, nodule, or sulci. The most frequent symptoms of speech pathology include the sound of your voice becoming scratchy, hoarseness excessive loudness, difficulty speaking effectively. Analytical or empirical approaches may be used to explore issues with speech or voice. Dysphonia, which is used to describe voice abnormalities that hinder a person from creating a sound utilizing the vocal organs, maybe a medical word for voice disorders. The endoscopic examination of the vocal folds as well as perceptual since auditory evaluation techniques as auditory evaluation methods are crucial for the treatment of dysphonia, as auditory evaluation methods are essential to the treatment of. This is why the Consensus Audio-Perceptual Evaluation Vocal together with other scales such as the GRBAS (grade roughness, breathiness asthenia strain) scales use ratings scales for measuring speech (including roughness of frequency the quality of breathiness and exhaustion, and stress) But, the CAPE-V is a measure of the degree of dysphonia in the form of an average and also the severity of breathiness, roughness and strain. But, as these techniques of evaluation are commonly used in clinical practice there are many aspects to be considered

in this test's analytical nature. There are many factors to take into consideration in determining treatment limitations, such as the experience of the clinician of the patient's dysphonia the form that the auditory perception rating system and the speech stimulation task the patient is performing. Following the failure of auditory evaluation, clinicians and researchers developed a more quantitative speech recognition metric to assess the amount of dysphonia in patients (Masud et al., 2012, p. 1015-1023). Acoustic examinations generate numerical values for severity of the disease as well as a treatment strategy, which are then released to other participants in the process of diagnosis and treatment. Assessments of the voice of patients typically rely on recorded vowel sustained recordings instead of continuous samples of speech. But even though it has been proven that continuous vowels are the most efficient method of getting an audio sample that captures various rhythms of expression, it is not representative of the way people speak (Wu et al., 2018). Similar to how the assessment of vocal consistency demands considering the timing of the start of a break, the end of a breath, and the breaks in a voice and vocal features, vocal changes regarding these three events cannot be fully conveyed by continuous vowel sounds like vowels. Thirdly, dysphonia signals are more evident when a person's voice is growing in conversation as opposed to sustained vowels and dysphonic people's gestures are more frequently expressed in a way that is different from those of non-dysphonic voices. If compared with the normal voice, the spasmodic dysphonia of the adductor is distinct by the amount of time that sounds are produced as compared with the normal voice. Acoustic correlates of an individual's voice have been developed because of the influence of the supra-segmental and phonetic structure of speech. It is not captured by the continuous vowel.

A specific set of skills is required to diagnose and treat voice difficulties properly. These skills include the ability to recognize certain abnormalities. An automated vocal pathology detection (VPD) system may assist clinicians in their work. The VPD technology works only to sustain vowels, meaning that the speech stream stays still for 6 to 9 seconds. However, residents should refrain from speaking for extended periods in everyday talks and

instead employ continuous expressiveness. A realistic VPD system must be conceptually capable of detecting pathology in successive sentences, which would imply that a practical VPD system would be capable of identifying pathology in continuous words that detect pathology in continuous phrases. According to these sources the method of voice recognition employed by VPD utilizes continuous speech, but it's not perfect.

Most of the time there are two primary reasons. The main reason is that we've not been able to obtain sufficient features for voice impairment. Therefore, the majority of the functions are supplied by speech processing or speaker recognition. There are a variety of types of voice problems that are divided into different categories. It is challenging to obtain details from vocal sounds and is reliable, efficient user-friendly and flexible characteristics that are capable of discrimination. Voice disorders are defined as issues with the volume, quality or pitch of the vocal sound that is generated by the larynx10. The causes of voice problems can result from a variety of causes that include mental health issues or trauma-related events or physical ailments, as well as illnesses, among others. Most of the time these speech issues aren't life-threatening and are relatively easy to correct. The most prevalent kinds of speech disorders are vocal abuse that occurs often. It is possible to look for long-lasting vocal signs, like polyps, nodules and cysts as well as an edoema (swelling) on the vocal folds in the event of a trauma to our vocals. The difficulty in speaking could be the result of a variety of conditions such as Parkinson's disease, endocrinological (hormonal) issues, and surgical procedures like surgery to remove thyroid glands or heart bypass. Based on the above review, we can think it's essential to examine the vocal condition at the time it develops. Numerous other VPD solutions have been mentioned in literature to now. Integration of VPD devices with intelligent healthcare is an emerging trend (Saarbruecken Voice Database, n.d.(.

AI machine learning (ML) deep learning cloud computing, edge computing, as well as new-generation wireless communications are developing to enhance the delivery of health services in a smart way. There are many types of detection systems for pathology are integrated into the

smart health framework (Ali et al., 2018, p. 19-28). These frameworks for smart health are becoming increasingly well-known because they improve the quality of our lives by allowing us to relax and feel at convenience. One can be diagnosed with a condition in their home, receive guidance from various physicians from all over the world and get rid of the pressure of obtaining an appointment at the hospital via the internet. The integration with technologies like Internet of Things (IoT) cloud computing, edge computing, as well as the 5G network allows for intelligent healthcare that is more accurate and less latency. The accuracy of healthcare systems has also improved due to advanced MI algorithms, like deep learning. In this research, we present the development of a speech pathology detection system within the context of a smart healthcare system that is smart. It makes use of two communication methods that include Speech signal, and the electroglottograph (EGG) signal. The convolutional neural model utilized to discover features in both of these modalities which is the way we build the system. By using the long-short-term memory model features are mixed. The tests (Daniel et al., 2019) use the Saarbrucken voice database, which the University of Saarbrucken developed.

In this research, the term Machine Intelligence is deliberately used instead of the more common term Artificial Intelligence (AI). While both concepts overlap, Machine Intelligence is employed here to emphasize the broader integration of intelligent computational methods, including but not limited to deep learning and data-driven decision-making. The adoption of this terminology also aims to distinguish this work from existing literature that frequently uses the term AI, thereby providing a unique perspective and avoiding redundancy in research titles.

II.Related Survey

There are several study papers on identifying voice pathology in the literature. We'll go through a few of them in more detail below. In the context of testing and voice evaluation, the measurement of voice quality is critical. Professional voice therapists working in facilities that have modern aerodynamic, acoustic, as well as vocal fold image equipment could utilize the auditory perceptual assessment method. Additionally, it is the case that

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

APA (American Psychological Association) gives a baseline of information about what degree of dysphonia that can be used to evaluate the progression of the patient's condition over time. According to APA, it is possible that, at the very least partially the cost-effectiveness as well as the short time commitment and comfort for patients of assessment of auditory perception of voice problems are a major factor in their effectiveness for treating vocal dysfunction. In addition, it evaluates how well speech sounds as well as the vocal strength by analysing certain auditory elements which can be discerned in the audio. There are many issues with regard to impartiality: The judges constantly disagree with each other, (ii) there are no quantitative measures or measurements, and (iii) there isn't a universal scale for measuring perceptual perception. The reasons are as follows They believe that scales by the senses can be a factor in the occurrence of errors and variations. Scales employed in clinical and research settings aren't always the most appropriate method for evaluating the voice quality characteristics.

On the other hand, the laryngoscopy review should not be utilized simply as a diagnostic tool in the evaluation of people who have laryngopharyngeal symptoms, given the poor Positive and negative predictive values for auditory and visual tests when coupled with the laryngoscopy examination. Numerous studies have demonstrated that there are only minor connections between tests that are instrumental (i.e., machines' measures of the quality of speech) as well as perceptual judgements of the quality of the voice (i.e., the perception of a listener's the quality of voice). Judges make use of a variety vocal stimuli, such as long vowels as well as flowing speech, when evaluating voice content. Some believe that flowing address is truer and more lifelike than continuous vowels (Guedes et al., 2019). Assessing for the presence of psychotic speech requires external rating systems with preset characteristics. According to Hammarberg(Chollet, 2017, p. 1800-1807), a frequent scale employed by Hirano Cummings are the GRBAS. The Committee for Testing of Phonatory Functions from the Japan Society of Logopedics and Phoniatrics created and approved the Phonetic Scales and Vocabulary Sets in 1969. GRBAS is the acronym as Grade in Dysphonia Roughness, Breathiness, Asthenia, and Strain. Based on the assessment of

parameters the results are: zero minor differences 1 moderate difference and 3 significant variations. Numerous risk factors can result in laryngeal disorders which can lead to the loss of voice over time (harmful chemicals, a lack of hygiene requirements in work that have high pressure on the voice and stress, etc.). Alterations to the larynx, such as preventing the vocal folds from vibrating properly, are possible on various levels. Tumors and some speech disorders, including laryngeal paralysis (semi- or complete), blame these improvements. The dysphonic voice or speaking will become hoarse as a result. One or more of the following may lead to a change in the voice: Adding noise or wind to the original voice; expansion of the vocal spectrum (or an increase in the volume of low-frequency sound); etc. are examples of verbal changes. Depending on the voice issue, any or all of these signs (symptoms) may be present. Diagnostic assistance systems based on acoustic voice analysis are being developed at numerous testing sites across the globe (Howard et al., 2017.(

Hadjitodirov et al. (Alhussein& Muhammad, 2018, p. 41034-41041) examined the diagnosis of laryngeal disease using sophisticated neural representations. Laryngeal illness diagnosis is projected to improve thanks significantly to their novel method, eliminating the most prevalent error in identifying patients with laryngeal diseases as ordinary speakers. Their so-called probability distribution map approach is built on the input vectors' probability density functions for normal and pathological data. The PDF of the unclear normal or abnormal issues was modelled using a template PDM. However, instead of only looking at the distance between two points, a method based on unique comparisons was used. Researchers in their research have developed a technique for screening laryngeal diseases. Nonlinear speech analysis approaches have been devised and substantially evaluated. In order to gather voice parameters for speech analysis, there has been consensus on how to do so. With the use of this approach, this research demonstrates that it is possible to restore the initial formant AM (amplitude modulation) modulation properties that were present before and after therapeutic voice therapy. According to the researchers, the extracted

function may be associated with both normal and abnormal patterns of vocal fold vibratory activity in the voice box.

Various nonlinear approaches were utilized to identify and classify automated vocal pathology identification and classification. This approach does not adequately describe machine nonlinearities since it does not operate with nonlinearities. It was shown that T parameters had a significant impact on the estimate of function regularity. Another approach used to estimate T is to utilize the baseline T was calculated from the embedding window and has a value of 1, which is used in a variety of approaches that apply the auto mutual information criterion, as well as in the usage of the embedding window. Nonlinear dynamic regression has been utilized in several investigations to investigate conventional evidence in healthy and dysphonic pediatric groups. Using perturbation approaches, jitter analysis may also identify symphonic populations, such as babies. A network-based transformation was used to demonstrate an alternative technique for pseudo periodic signals. New methods for pathological assessment were proposed to help identify moderate and pathological subjects. In(Alshehri& Muhammad, 2021, p. 3660-3678), dynamic networks were referred to as novel notions in a new approach for removing features. Visually differentiating between healthy persons and those suffering from the disease is easier using this method.

A variety of binary rates and compressed MP3 voice samples were employed in a paper (Muhammad et al., 2019, p. 44-49). Gaussian Mixture Model and Support Vector Machine are used as classification tools in this research. There is a significant relationship between MFCCs and the accuracy of classification according to Wang and co. (Muhammad et al., 2021, p. 603-610). Other methods could help in improving the process of classifying. Researchers studied two different voice classification systems in this study including the Gaussian Mixture Model classifier and the Super Vector Machine. A whopping 96.1 percent of sustained vowel phonation was shown to be feasible. In their study, Jang et al. tested 99 individuals with vocal fold polyps, cysts, and nodules on various pitch detection algorithms (PDAs). The researchers then concluded that PDAs were

sufficient. Pitch errors were shown to be more frequent when the vocal topic was more chaotic and aperiodic. Researchers Gomez-Velda et al. used biomechanical data as features in another investigation. This characteristic was derived from a vocal sound sample and may be used to estimate the displacement of the vocal folds without the need for invasive methods. The study included 52 patients who had polyps, nodules, recurrent laryngitis, and Reinke'sedoema due to surgery for vocal fold polyps.

III.Methods and Materials

Database used

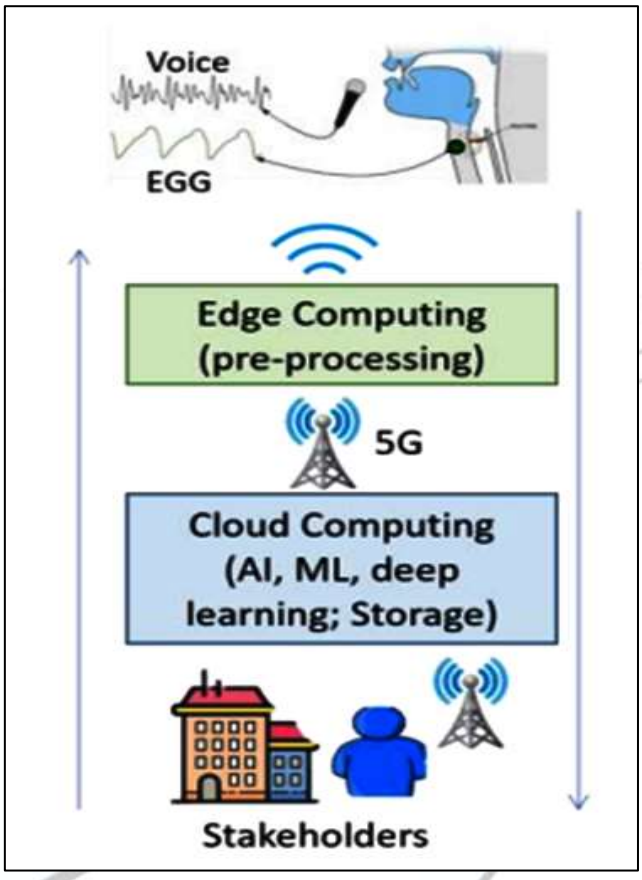
More than 2000 people and 71 distinct voice disorders may be found in the SVD database (Daniel et al., 2019), which is commonly used in speech pathology research, is a good example. During the system's training, testing, and verification, the system was fed with persistent /a/ speech signals and EGG signals. There were 842 groups in each training sample, 791 problematic groups (corresponding to 60% of the entire sample), and 281 healthy groups (equivalent to 20% of the total sample) in each verification sample, for a total of 842 groups per verification sample. The experiments made use of a variety of speakers ranging in age from 15 to sixty-five years old.

Proposed model

In this article it is presented that a VPD system is discussed in the framework of a smart health framework. Smart healthcare architectures include the Internet of Things, deep learning edge computing, cloud computing, as well as 5G connectivity. In figure 1 an intelligent healthcare system that detects speech pathology is a prime illustration of this type of technology. Within this framework Internet of Things devices like microphones or EGG devices are utilized to collect the signals of an individual. When it comes to obtaining spectrums, these signals are then sent to edge computing for processing prior to being processed. The spectrums are then uploaded to the cloud where they are kept and processed by AI/ML/deep learning servers. The final decision is then made available to all stakeholders as well as the user by using 5G technology.

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

Figure 1: The VPD system to improve from a smart healthcare framework



VPD System

EGG and speech signals are used in figure 1 to demonstrate the proposed VPD system. The fusion follows separate processing of the signs in a subsequent step. Voice signals are captured by a microphone and EGG signals by an EGG device. Short-time framing (30-millisecond frames with ten-millisecond overlap), hammer windowing, and the short-time Fourier transforms are used to build spectrograms of the data, after they have been trimmed to remove bias. Additionally, we make use of Mel-spectrograms that are created by applying Mel band-pass filters that are scale-spaced on the data used for investigation (36 Mel filters). The high-order harmonic distortion of the spectrogram is decreased because of the processing of the

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

voice samples prior being converted to STFT. Amount of feature mapping data decreased when the sampling rate is decreased by 16kHz. This means that since there's less data handle, the training process can be completed faster. The emphasis is pre-set on the frame to increase the high-frequency clarity of the speech, The workflow of the proposed Voice Pathology Detection system is illustrated in Figure 2.

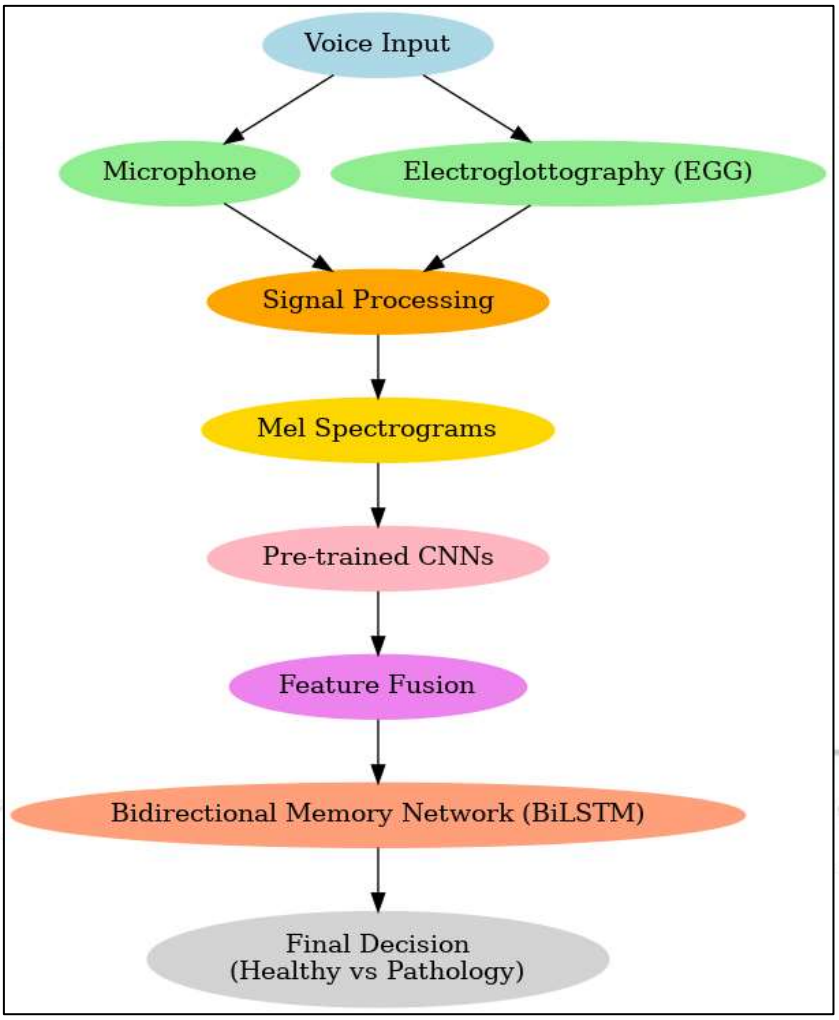


Figure 2: VPD System in Smart Healthcare Framework

JOBS مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

This flowchart shows the steps of the proposed Voice Pathology Detection (VPD) system, starting from patient voice input → capturing signals via microphone and EGG → signal processing → CNN models → feature fusion → BiLSTM processing → final decision (Healthy vs. Pathology.(

The spectrograms are put into a CNN model that has already been trained. In our research, we use the ResNet50, Xception, and MobileNet networks. Given the limited amount of data we have to deal with, pre-trained models are the most efficient method of getting the system up and running. Table 1 presents a high-level summary of the three CNN models under consideration.As a result, MobileNet is less complicated to use for real-time applications than ResNet50 and ResNet50, allowing it to be implemented more quickly.

Nevertheless, a CNN with a high number of parameters may be employed for quick processing because of the continual progress in processing speed. The input size of ResNet50 and MobileNet is 224224, whereas the input size of the Xception is 299299. The spectrograms and Mel-spectrograms are scaled following the input size. A typical person's voice and EGG signals are shown in the top row of Table 1, together with accompanying spectrograms and Melspectrograms. The LSTM model is made up of LSTM units [40]. The LSTM unit has three gates: input, forget, and output. LSTM gates allow the transmission of both real-time and hidden history data about a given point in time. For the purpose of calculating the significance of the gates, three completely linked layers that utilise the sigmoid function to calculate the values of the input, forget, and output gates are used. In order to build an LSTM layer, LSTM units are stacked one on top of the other on top of one another. It is possible to mix these LSTM layers to produce either bidirectional or unidirectional LSTM networks. The forward and backward time directions of two layers of a bidirectional LSTM are related in both directions, and the layers operate in both directions (BiLSTM). It is feasible to learn time-dependent correlations via the use of successive layers in this manner. Each BiLSTM layer contains a total of 256 stacked LSTM units, which is the maximum number possible. SoftMax is utilized as the BiLSTM layer to determine embedded patterns. This is the place where SoftMax plays

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

a role. First it is used to train the CNN model to identify the features, and then freeze it to preserve the features that were discovered. These features are then fed to the BiLSTM model to aid in the extraction of temporal features as explained previously. If a dropout that is 50 percent is applied prior to the fully linked layer is completed it is considered to be functional. This loss function rests upon the measure of cross-entropy

CNN model	Xception	ResNet50	MobileNet
Input size	299 299	224 224	224 224
Parameters	22,9 Million	25,6 Million	4,2 Million
Size (MB)	88	96	16
Depth	126	—	88
Layers number	71	50	28

Table 1: general information of (Xception, ResNet50, MobileNet) CNN models

IV. Experimental Evaluation

A large number of experiments were conducted in order to verify the concept of VPD system. The method was evaluated against other similar methods that were found in scientific research. To evaluate the efficacy of the different strategies under review the four performance criteria of precision, accuracy, recall and F1-score were utilized. The term "precision" refers to the percentage of all samples that are included in the set that could be accurately predicted to be taken into the collection. It is a measure which measures the effectiveness of a model by reliably recognizing negative data. The precision rate is measured as percentage of negative data identified. It's the proportion of test results that were expected to be positive. This is also known as recall. A higher rate of recall indicates that this model has been proven to be more reliable in identifying positive samples, which is crucial in conducting a study. A higher F1 score suggests that the individual has more ability to classify information.

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

As illustrated in Figure 3, spectrograms of audio samples recorded under various conditions are shown. The top row represents signals captured by different microphones in an office room environment, while the bottom row demonstrates the effect of different recording environments (silent room, cafeteria, office room) using the Yamaha Mixer microphone. These spectrograms highlight the variability in signal characteristics caused by microphone type and environmental conditions, which is a crucial factor to consider in audio-based pathology detection systems.

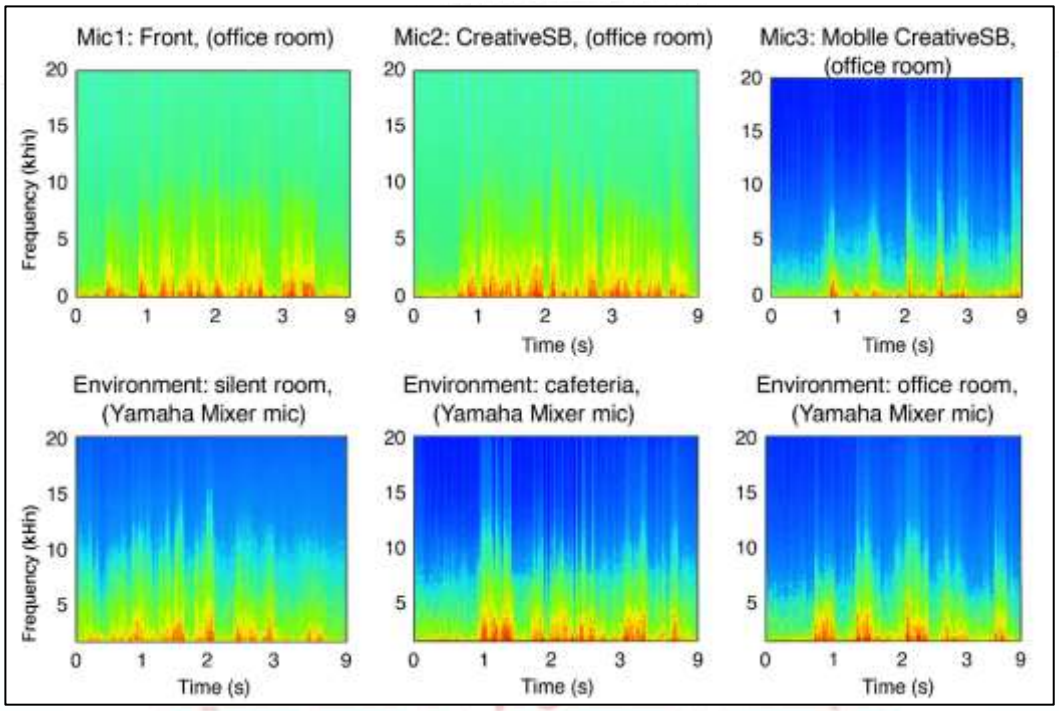


Figure 3: Spectrograms of audio samples recorded with different microphones and environments

An Adam optimizer was used to optimize the model's parameter values. Training epochs were completed 200 times at a learning rate of 10-4. The batch size was 32, and the learning rate was 10-4; these were the experiment's parameters. Using an Adam optimizer, we can achieve a learning rate of 10-4 while maintaining batch sizes of 32 and 100 training epochs, respectively, in our optimization strategy.

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

The comparative performance metrics of voice, EEG, and the proposed fusion method are shown in Figure 4.

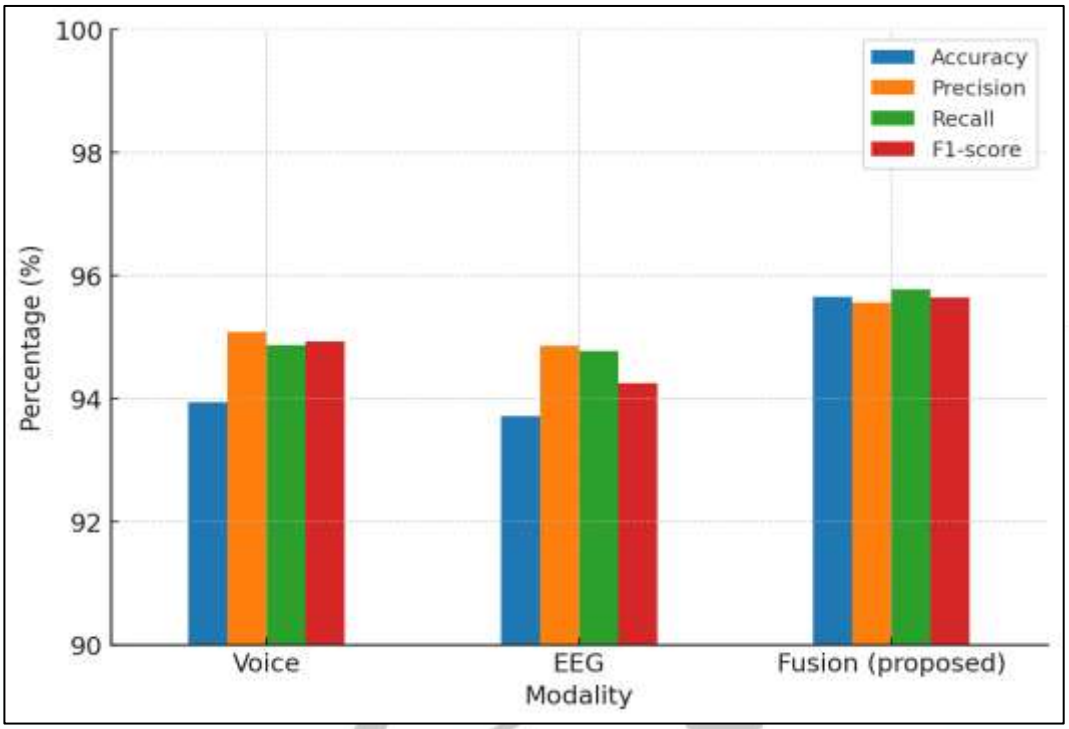


Figure 4: Performance Metrics of Different Models

This bar chart compares Accuracy, Precision, Recall, and F1-score between three approaches: voice only, EEG only, and the proposed fusion system.

Modality	Accuracy	Precision	Recall	F1-score
Voice	93.94	95.08	94.87	94.93
EEG	93.71	94.86	94.77	94.25
Fusion (proposed)	95.65	95.56	95.78	95.64

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

Table 2: performance metrics of different models, including proposed model

In the table, results of the four-performance metrics obtained through the implementation of this recommended method are as following the single-modality (voice) and single-modality as well as bi-modality systems were evaluated on their performance within this study (the system that was proposed). It is evident from this table that the system proposed outperformed single modality by a significant margin. This results in an enhancement in the overall performance of a VPD system as a consequence of the combination of speech and EGG signals. As illustrated in Table 2 and Figure 5, the proposed fusion-based system significantly outperforms single-modality approaches (voice-only or EGG-only) across all performance metrics, including accuracy, precision, recall, and F1-score. This demonstrates the effectiveness of combining voice and EGG signals in improving voice pathology detection performance. The Xception model was used in order to get the arrangement shown in Table 2.

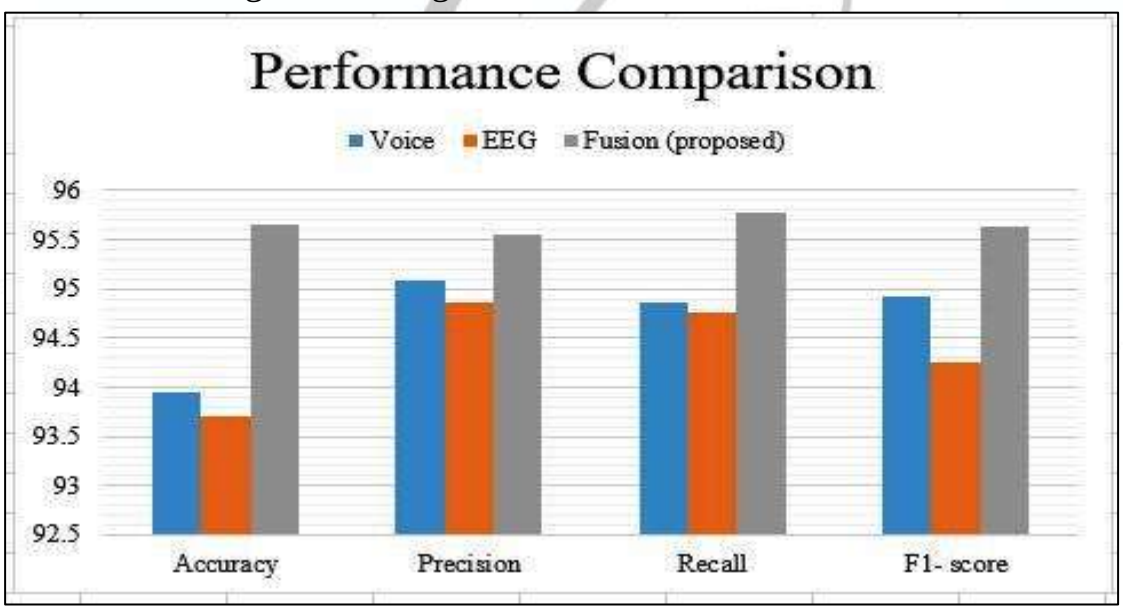


Figure 5: Performance comparison of different models

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

Three distinct pre-trained CNN models were examined for their performance in the proposed system. The accuracy results of different CNN models used in the proposed framework are presented in Figure 6.

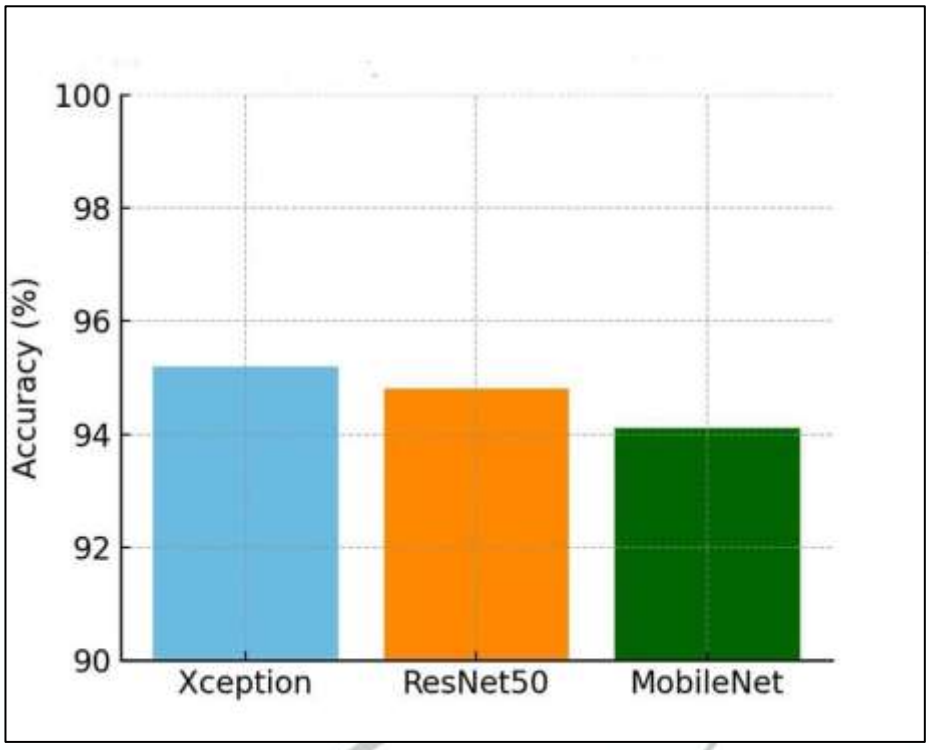


Figure 6: Accuracy Comparison of CNN Models

This bar chart presents a comparison of the accuracy of three convolutional neural network (CNN) models: Xception, ResNet50, and MobileNet. As shown in Figure 6, the Xception model achieved the highest performance, outperforming the other two in overall accuracy. Although MobileNet has significantly fewer parameters, its performance remained competitive and did not fall far behind the other models.

We compared the suggested system to other relevant designs utilizing the same database as the proposed system. It should be mentioned that the integration of speech and EGG signals is an uncommon the frequency of occurrences in the scientific literature. Table 3 presents a comparison of the efficiency of various methods.

JOBS مجلة العلوم الأساسية Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ

System	Accuracy (%)
Proposed (Fusion)	95.65
Voice only	93.94
EEG only	93.71
Other AI Models	92.5

Table 3: Comparison with Other Methods

This table compares the proposed system with other existing methods, showing that the proposed method achieved the best accuracy.

The results show in this table that the proposed method was superior to all the other systems tested. There are a variety of other artificial intelligence systems and IoT-based smart healthcare systems (Hossain & Muhammad, 2018, p. 2399-2406), however research on the application for speech pathology recognition in healthcare technology is confined. Because voice pathology significantly impacts instructors, a smart classroom setting may include the VPD system to ensure long-term teaching success (Saarbrucken Voice Database, n.d.).

ImplementationEnvironment

The experiments were implemented using Python 3.8 with TensorFlow 2.9 and Keras 2.8 for model development, along with NumPy, SciPy, Librosa, and Scikit-learn for signal processing and feature extraction. The system was executed on a workstation equipped with an Intel Core i7-10750H CPU @ 2.60GHz, 16GB RAM, and an NVIDIA GeForce GTX 1660 Ti GPU, running on Windows 10 (64-bit). The Saarbrucken Voice Database was employed and the dataset was divided into 80% training and 20% testing subsets for evaluation.

I.Discussion

The conducted experiments aimed at validating the efficacy of the VPD (Voice Pathology Detection) system, comparing it against existing methods in scientific research. Key performance metrics including precision,

accuracy, recall, and F1-score were employed to assess different strategies. Precision, as a measure of a model's effectiveness in recognizing negative data, and recall, which indicates the model's reliability in identifying positive samples, were crucial factors in evaluating the system's performance. A higher F1-score reflects the system's enhanced ability to classify information accurately. Utilizing an Adam optimizer, the model's parameters were optimized, with specific parameters such as training epochs and batch size meticulously chosen to ensure effective optimization. Notably, the proposed system outperformed single-modality approaches, particularly with the fusion of speech and EGG signals, as demonstrated by the significant improvements in performance metrics. The choice of the Xception model further contributed to the system's superior performance, as illustrated in performance comparison figures. Comparison with existing methods highlighted the rarity of integrating speech and EGG signals in scientific literature, underscoring the novelty and effectiveness of the proposed approach. These findings suggest promising applications for the VPD system, particularly in healthcare and educational settings, where speech pathology recognition can significantly impact outcomes, potentially improving long-term teaching success in smart classroom environments.

II. Conclusion and future Enhancement

This research offered a pathological speech identification approach that relied on a bimodal input to function properly. The procedure accepts as inputs the speech signal, as well as the EGG signal with the proposed VPD system, calls are transformed into spectrograms, which are then input into a CNN model for further analysis. The next phase is integrating the features gathered from the two modalities and incorporating them into the BiLSTM model as a whole. The experimental data demonstrated that the proposed strategy attained accuracy, precision, and recall rates more than 95 percent in each of the three categories tested in the experiments. Furthermore, the system outperformed other systems of a similar kind. It has been proven that the accuracy of bimodal inputs outperformed that of single inputs. By experimenting with the VPD system, we will investigate the impact of signal transmission through a network in the future.

III. References

1. Muhammad, Ghulam., Rahman, Sheikh K. M. Mohd., Alelaiwi, Abdulhameed., & Alamri, Atif. *"Smart health solution integrating IoT and cloud: A case study of voice pathology monitoring."* IEEE Communications Magazine, Vol. 55, Issue 1, pp. 69–73, January 2017.

2. Ali, Zulfiqar., Elamvazuthi, Irraivan., Alsulaiman, Mansour., & Muhammad, Ghulam. *"Automatic voice pathology detection with running speech by using estimation of auditory spectrum and cepstral coefficients based on the all-pole model."* Journal of Voice, Vol. 30, Issue 6, pp. 757.e7–757.e19, November 2016.

3. Al-nasheri, Abdullah., Muhammad, Ghulam., Alshehri, Fahad., Amin, Syed Umar., & Hossain, Md. Shamim. *"Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions."* IEEE Access, Vol. 6, Issue 1, pp. 6961–6974, December 2018.

4. Ali, Zulfiqar., Khan, Rehan., & Javed, Faizan. *"Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals."* Journal of Medical Systems, Vol. 40, Issue 20, pp. 1–10, 2016.

5. Masud, Mohammad., Hossain, Md. Shamim., & Alamri, Atif. *"Data interoperability and multimedia content management in e-health systems."* IEEE Transactions on Information Technology in Biomedicine, Vol. 16, Issue 6, pp. 1015–1023, November 2012.

6. Hossain, Md. Shamim. *"Cloud-supported cyber–physical localization framework for patients monitoring."* IEEE Systems Journal, Vol. 11, Issue 1, pp. 118–127, March 2017.

7. Muhammad, Ghulam., Almalki, Mohammed., & Al-Hussein, Mohammed. *"Automatic voice pathology detection and classification using vocal tract area irregularity."* Biocybernetics and Biomedical Engineering, Vol. 36, Issue 2, pp. 309–317, 2016.

8. Hossain, Md. Shamim., Muhammad, Ghulam., & Alamri, Atif. *"Smart healthcare monitoring: A voice pathology detection paradigm for smart cities."* Multimedia Systems, Vol. 25, Issue 5, pp. 565–575, 2019.

9. Amin, Syed Umar., Hossain, Md. Shamim., & Muhammad, Ghulam. *"Cognitive smart healthcare for pathology detection and monitoring."* IEEE Access, Vol. 7, Issue 1, pp. 10745–10753, December 2019.

10. Hossain, Md. Shamim., Alamri, Atif., & Muhammad, Ghulam. *"Audio-visual emotion-aware cloud gaming framework."* IEEE Transactions on Circuits and Systems for Video Technology, Vol. 25, Issue 12, pp. 2105–2118, December 2015.

11. Hossain, Md. Shamim., & Muhammad, Ghulam. *"Emotion-aware connected healthcare big data towards 5G."* IEEE Internet of Things Journal, Vol. 5, Issue 4, pp. 2399–2406, August 2018.

12. Saarbruecken Voice Database. *"Database help and documentation."* Retrieved from http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4, April 2021.
13. Ali, Zulfiqar., Rehman, Abdur., & Rauf, Hafeez Tariq. *"An intelligent healthcare system for detection and classification to discriminate vocal fold disorders."* Future Generation Computer Systems, Vol. 85, pp. 19–28, August 2018.
14. Al-nasheri, Abdullah., Alshamrani, Hani., & Khan, Rehan. *"Investigation of voice pathology detection and classification on different frequency regions using correlation functions."* Journal of Voice, Vol. 31, Issue 1, pp. 3–15, 2017.
15. Pavol, Harar., & Hossain, Md. Shamim. *"Voice pathology detection using deep learning: A preliminary study."* In International Conference and Workshop on Bioinspired Intelligence (IWOBI). IEEE, July 2017.
16. Daniel, Korzekwa., Albrecht, Robert., & Klein, Jan. *"Interpretable deep learning model for the detection and reconstruction of dysarthric speech."* In International Speech Communication Association (ISCA), Graz, Austria, September 2019.
17. Wu, Hao., Zhang, Xinyu., & Sun, Jian. *"Convolutional neural networks for pathological voice detection."* In 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 1–4, Hawaii, United States, July 2018.
18. Mohammed, Mahmood A., Abdulkareem, Khaled H., & Taha, Mustafa H. N. *"Voice pathology detection and classification using convolutional neural network model."* Applied Sciences, Vol. 10, Issue 3723, pp. 1–12, 2020.
19. Guedes, Vitor., Figueiredo, Carlos., & Sousa, Joao. *"Transfer learning with AudioSet to voice pathologies identification in continuous speech."* In International Conference on ENTERprise Information Systems (CENTERIS), Sousse, Tunisia, October 2019.
20. Hossain, Md. Shamim., & Muhammad, Ghulam. *"Deep learning-based pathology detection for smart connected healthcare."* IEEE Network, Vol. 34, Issue 6, pp. 120–125, November 2020.
21. He, Kaiming., Zhang, Xiangyu., Ren, Shaoqing., & Sun, Jian. *"Deep residual learning for image recognition."* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, Las Vegas, NV, June 2016.
22. Chollet, Francois. *"Xception: Deep learning with depthwise separable convolutions."* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1800–1807, Honolulu, HI, July 2017.
23. Howard, Andrew G., Zhu, Menglong., Chen, Bo., Kalenichenko, Dmitry., Wang, Weijun., Weyand, Tobias., ... & Adam, Hartwig. *"MobileNets: Efficient convolutional neural networks for mobile vision applications."* arXiv preprint arXiv:1704.04861, 2017.
24. Alhussein, Mohammed., & Muhammad, Ghulam. *"Voice pathology detection using deep learning on mobile healthcare framework."* IEEE Access, Vol. 6, pp. 41034–41041, December 2018.

25. Alshehri, Fahad., & Muhammad, Ghulam. *"A comprehensive survey of the Internet of Things (IoT) and AI-based smart healthcare."* IEEE Access, Vol. 9, pp. 3660–3678, January 2021.
26. Muhammad, Ghulam., Alhamid, Mohammed F., & Long, Xiaofei. *"Computing and processing on the edge: Smart pathology detection for connected healthcare."* IEEE Network, Vol. 33, Issue 6, pp. 44–49, November–December 2019.
27. Muhammad, Ghulam., Hossain, Md. Shamim., & Kumar, Neeraj. *"EEG-based pathology detection for home health monitoring."* IEEE Journal on Selected Areas in Communications, Vol. 39, Issue 2, pp. 603–610, February 2021.

JOBS
مجلة العلوم الأساسية
Journal of Basic Science
Print -ISSN 2306-5249
Online-ISSN 2791-3279
العدد الواحد والثلاثون
٢٠٢٥م /١٤٤٧هـ