وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم علوم الحاسوب

نظام للتعـــــرف علـى المـــشاعر من الكلام للغات متعددة باستخدام التعلم الآلي

رسالة مقدمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
كجزء من متطلبات نيل شهادة ماجستير علوم في
علوم الحاسوب

من قبل

علي عبد الوهاب يحيى الصفار

بإشراف

أ.د. فوزية محمود رمو

1445 هـ

2023 م

الملخص

يعد الكلام من أهم أدوات التواصل بين الأفراد وتحقيق التفاهم والعيش بسلام , ويؤدي الكلام دور أساسي لنقل مشاعر الانسان بناءً على ما يقوله الآخرون , إن الغرض من بناء هذا النظام هو جعل الآلة تتعرف على المشاعر البشرية في ضوء الكلام ، ليتم استثمار هذا التفاعل في تطبيقات مفيدة تخدم البشر.

تم بناء نظام حاسوبي ذكي لتمييز خمسة مشاعر(غضب , سعادة , طبيعية , حزن , تفاجئ) من الكلام معتمد على مجموعة البيانات متعددة (Multi Dataset) أطلق عليه اسم (Speech Emotion Recognition System) و اختصاراً (SERSYS) .

يتكون هذا النظام من نموذجين أساسيين، كل نموذج يقسم إلى نموذجين فرعيين , النموذج الأول يعتمد على مبدأ الشبكة العصبية العميقة يطلق عليه اسم (SERDL) أختصاراً لـ (Speech Emotion Recognition Deep Learning).النموذج الثاني يعتمد على مبدأ نقل التعلم والشبكات المدربة مسبقا (Speech Emotion Recognition Transfer Learning) يطلق عليه اختصارا (SERTL) .

يتكون نظام (SERSYS) المقترح من عدة مراحل، في المرحلة الأولى يتم تهيئة مجموعات البيانات(Dataset) لملفات الصوت حيث تم استخدام مجموعة بيانات (TESS) خاصة بالمشاعر لملفات صوتية باللغة الانكليزية ومجموعات بيانات خاصة بالمشاعر للغة الفرنسية سميت ب (caFE) و للغة الإيطالية ب (EMOVO) تم دمجهم في مجموعة بيانات واحدة باسم (Multi dataset), تليها المرحلة الثانية مرحلة استخلاص الصفات (Features Extraction) باستخدام خوارزمية معاملات درجة النغم (MFCC) أختصاراً لـ(Mel Frequency Cepstral Coefficients), تم خزن المعاملات الناتجة بصيغة صورة مخطط طيفي وبمعدل عينة (22050 Hz) حيث تم استخلاص الصفات بطريقة MFCC والأرقام 26 و39 تمثل عدد الميزات المستخلصة من كل اطار صوتي في الكلام .

المرحلة الثالثة هي مرحلة التعرف على المشاعر باستخدام طريقتين ذكائيتين , **الأولى** في استخدام الشبكة العصبية العميقة (Deep Neural Network) متمثلة بالنموذج (SERDL Model) ,الثانية باستخدام الشبكات المدربة مسبقا(ResNet50,VGG16) مع

مصنفات التعلم الآلي مثل الة متجة الدعم(Support Vector Machine) والغابات العشوائية (Random Forest) متمثلة بالنموذج (SERTL) .

تم الحصول على أفضل النتائج في التعرف على المشاعر في اللغات المنفردة وكانت الدقة (99 %) للكلام باللغة الإنكليزية عند استخدام 39 ميزة (MFCC) كإدخالآت للنموذجين (SERTL , SERDL) و(91 %) للكلام باللغة الفرنسية و (88%) للغة الايطالية مع استخدام 26 ميزة (MFCC) كإدخال للنموذج (SERTL).

في حالة تصنيف فئات الشعور للغات متعددة(Multi languages) , تم الحصول على أفضل دقة تصنيف بلغت (94%) باستخدام نموذج (SERTL_ML) مع شبكة (ResNet50) ومصنف (RF) وباستخدام عدد معاملات درجة نغم (MFCC=26).

وبالاعتماد على النظام (SERSYS) قمنا بانشاء نظام آخر باسم (CHESS_SER SYSTEM) والذي يقوم بكشف حالتين عاطفيتين باللغة العربية وهي الطبيعي(وتمثل حالة الهدوء والراحة) و غير الطبيعي (وتمثل حالة الغضب أو الحزن أو عدم الراحة) , تم تطبيقها على 5 لاعبين من أبطال جامعة نينوى في لعبة الشطرنج وحصلنا على دقة تصنيف بلغت (98%) باستخدام نموذج (SERDL) ودقة (89%) باستخدام نموذج (SERTL).

Ministry of Higher Education and

Scientific Research

University of Mosul

College of Computer Science and Mathematics

A System for Emotion Recognition From Speech for Multi Languages Using Machine Learning

A Thesis Submitted to the Council of the College of Computer Science and Mathematics University of Mosul as a Partial Fulfillment of Requirements for the Degree of Master of Science in Computer Science

By

Ali Abdulwahhab Y. Alsaffar

Supervised by

Prof. Dr. Fawziya Mahmood Ramo

2023 A.D. **1445 A.H.**

Abstract

Speech is a crucial tool for effective communication, understanding, and peaceful coexistence among individuals. It enables people to express their emotions and convey their thoughts effectively. The aim of developing this system is to enable machines to recognize human emotions through speech, thereby facilitating the development of useful applications that serve humanity.

To achieve this goal, a sophisticated computer system was created that can identify five different emotions - anger, happiness, normalcy, sadness, and surprise - from speech using a multi-dataset approach. This system is known as the Speech Emotion Recognition System (SERSYS).

The SERSYS comprises two main models, each of which has two sub-models. The first model is based on the principles of deep neural networks and is called Speech Emotion Recognition_Deep Learning (SERDL). The second model uses transfer learning and pre-trained networks and is known as Speech Emotion Recognition_Transfer Learning (SERTL). Together, these models enable the machine to accurately recognize human emotions from speech data.

Overall, the development of SERSYS represents a significant step forward in the field of artificial intelligence and has enormous potential for improving human-machine interactions in various applications.

The SERSYS system is comprised of several stages. Firstly, datasets are created for voice files using TESS for English language and caFE for For French and for Italian languages EMOVO is used, which are merged into a Multi dataset. Secondly, the MFCC algorithm is used for feature extraction, resulting in spectral chart images stored at a sample rate of 22050 Hz. The third stage involves recognizing speech emotions using two intelligent methods: SERDL Model and SERTL Model with pre-trained networks (ResNet50, VGG16) and machine learning classifiers (Support Vector Machine), Decision Trees, and Random Forests. The best results were achieved in single-language emotion recognition with accuracy rates of 99% for English speech using 39 MFCC features as inputs for SERDL Model, 91% for French speech and 88% for Italian speech using 26 MFCC features as inputs for SERTL Model. For multi-

language emotion classification, the best accuracy rate of 94% was achieved using SERTL_ML Model with ResNet50 network and RF classifier with 26 MFCC coefficients.

We have developed a new system called CHESS_SER SYSTEM, using the SERSYS framework. This system is capable of detecting two emotional states in Arabic language - normal (indicating calmness and comfort) and abnormal (indicating anger, sadness or discomfort). We tested this system on 5 players from Nineveh University who were playing chess. Using the SERDL model, we achieved a classification accuracy of 98%, while the SERTL model gave an accuracy of 89%.