التعلم العميق للنمذجة الصوتية في انظمة التعرف على الكلام باللغة العربية

مقدمة من

إنجي رجائي راضي عبدالمقصود

مدرس مساعد - قسم العلوم الأساسية – كلية الحاسبات والمعلومات – جامعة الفيوم

رسالة مقدمة إلي قسم الفيزياء، كلية العلوم، جامعة الفيوم

كجزء من متطلبات الحصول علي درجة الدكتوراه الفلسفية

في

فيزياء الجوامد التجريبية

تحت إشراف

أ. د. عرفه صبرى جمعه حسن

قسم الفيزياء
كلية العلوم
جامعة الفيوم

أ. د. نبيلة محمد حسن

قسم العلوم الأساسية
كلية الحاسبات والمعلومات
جامعة الفيوم

أ. د. محمد هشام فاروق

قسم الرياضيات والفزيقا الهندسية
كلية الهندسة
جامعة القاهرة

كلية العلوم
جامعة الفيوم
٢٠٢١

ملخص الرسالة

تنقسم هذه الرسالة إلي جزئين، الجزء الاول ركزعلى التعرف علي الاصوات العربي باستخدام كلمات منفصلة. تم استخدام تقنيات مختلفة أثناء استخراج خصائص الصوت مثل MFSC، GFCC بمشتقاتها من الدرجة الأولى والثانية. تُستخدم الشبكة العصبية التلافيفية (CNN) لأداء تعلم الميزات والتصنيف. حققت CNN أداءً جيد في التعرف التلقائي على الكلام (ASR). يعد الاتصال المحلي ومشاركة الوزن والتجميع من الخصائص الرئيسية لشبكات CNN التي لديها القدرة على تحسين ASR . لقد تم اختبار نموذج CNN علي قاعدة البيانات لبعض مقاطع اللغة العربية. تم تعزير قاعدة البيانات المستخدمة من خلال تطبيق بعض التحويلات مثل تغيير درجة الصوت والسرعة والنطاق الديناميكي وإضافة الضوضاء علي المقاطع. وجد أن أقصى دقة تم الحصول عليها عند استخدام GFCC مع CNN هي ٩٩.٧٧٪. تم مقارنة نتائج هذا العمل بنتائج الابحاث السابقة ولوحظ أن شبكة CNN حققت أداءً أفضل في ASR.

الجزء الثاني يقدم نظام جديد للتعرف التلقائي على العاطفة باستخدام التعرف التلقائي على الكلام (ASR) باستخدام الشبكة العصبية التلافيفية (1-D CNN) لكلام العربي. نقدم نتائج للتعرف على المشاعر الثلاثة: السعادة والغضب والمفاجأة. يتم استخدام مجموعة بيانات الصوت العربي الطبيعي (ANAD). يتم استخراج خمسة وعشرين ميزة (LLDs) من الإشارات الصوتية. يتم فحص مجموعة مختلفة من الميزات المستخرجة. كما تم فحص تأثير استخدام تقنية تحليل المكونات الرئيسية (PCA) لتقليل الأبعاد. لمرحلة التصنيف ، يتم استخدام 1-D CNN و DFFNN. أيضًا ، تتم التعامل مع مشكلة عدم تساوي العينات في كل مجموعة من البيانات باستخدام تقنية (B-SMOTE). اتضح من النتائج أن أفضل دقة يتم الحصول عليها عند تطبيق استخدام جميع الميزات المستخرجة مع ال CNN هي ٩٩.٠٥٪. أيضًا ، تبلغ الدقة 98.92٪ عند استخدام LSF و MFCC. لا تختلف هذه النتيجة كثيرًا عن دقة استخدام جميع المميزات المستخرجة. ويلاحظ أن الدقة 98.11٪ عند استخدام خصائص LSF مما يدل على أنها من السمات السائدة. أظهرت النتائج المتحصل عليها تحسنا مقارنة بالدراسات السابقة.