عنوان البـــــحث باللغة العربية:

المشفر التلقائي المتغير ثنائي النسق للتعرف على الكلام من الإشارات السمعية و البصرية

ا . د. شيرين علي محمد طايع

ملخص البـــــحث باللغة العربية :

الانصهار متعدد الوسائط هو فكرة الجمع بين المعلومات في تمثيل مشترك متعدد الأشكال. الهدف من الاندماج متعدد الوسائط هو تحسين دقة النتائج من التصنيف أو مهام الانحدار. تقترح هذه الورقة ترميز تلقائي متغير ثنائي النسق (BiVAE) لدمج الميزات السمعية والبصرية. الاعتماد على الإشارات السمعية والبصرية في مهمة التعرف علي الخطاب تزيد من دقة التعرف، خاصة عند تلف إشارة الصوت. تم تدريب نموذج BiVAE والتحقق من صحته على مجموعة بيانات "CUAVE". ثلاثة مصنفات قاموا بتقييم الميزات السمعية والبصرية المدمجة: الذاكرة طويلة-قصيرة المدى "Long-short Term Memory" ، والشبكة العصبية العميقة "Deep Neural Network" ، وآلة المتجهات الداعمة "Support Vector Machine". تتضمن التجربة تقييم الميزات السمعية والبصرية المدمجة في حالة ما إذا كانت هناك طريقتان متاحتان أو أن هناك طريقة واحدة فقط متاحة (أي ، عبر إشارات الوسائط). أظهرت النتائج التجريبية تفوق النموذج المقترح (BiVAE) لدمج الميزات السمعية والبصرية على أحدث النماذج من خلال متوسط فرق الدقة ≃ ٣.٢٨٪ و ١٣.٢٨٪ لإشارة الصوت النظيفة ولإشارة الصوت التي تحتوي علي الضوضاء ، على التوالي. بالإضافة إلى ذلك ، يتفوق BiVAE على أحدث النماذج في حالة إشارات الوسائط المتقاطعة بفارق دقة ≃ ٢.٧٩٪ عندما تكون الإشارة الصوتية الوحيدة المتاحة و ١.٨٨٪ عندما تكون إشارة الفيديو الوحيدة المتاحة. علاوة على ذلك ، تحقق Support Vector Machine (SVM)أفضل دقة تمييز مقارنة بالمصنفات الأخرى.