اختيار اللاعبين الرياضيين باستخدام التقنيات الذكائية

م.م. محمد نايف عبد كنه
mohammed_naif85@nan.epedu.gov.iq

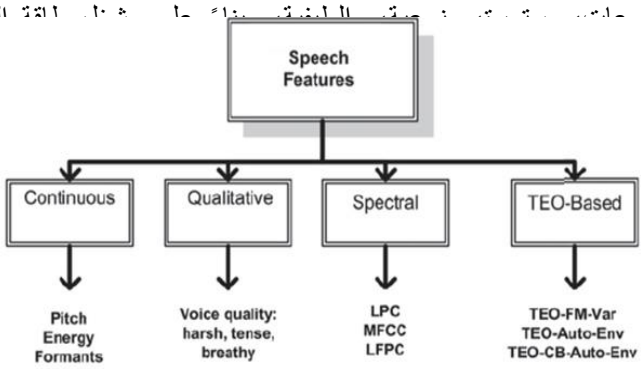
المســـتخلص: الكلام هو أحد أكثر أشكال التعبير الطبيعية. مع تقدم التكنولوجيا، أصبح التفاعل بين البشر والآلات أكثر تعقيدا. أظهرت الأبحاث أنه يمكن اكتشاف العواطف ليس فقط من خلال تعبيرات الوجه ولكن أيضًا من خلال الأصوات. مثل ملامح الوجه، يمكن للخصائص السمعية مثل طبقة الصوت وشدتها أن توفر معلومات قيمة حول الحالة العاطفية للشخص. وتشير الدراسات إلى أنه من الممكن بناء أنظمة ذكية تتعرف على الحالات العاطفية للأفراد من خلال أصواتهم، وتميز بين الغضب والحزن والسعادة والخوف والحالة الطبيعية وغيرها. ويمكن استخدام هذه الأنظمة في تطبيقات مختلفة مثل أنظمة سلامة السائق وأنظمة خدمة العملاء. نهدف في هذه الدراسة إلى تطبيق أنظمة التعرف على المشاعر في المجال الرياضي. تم تطوير نظام يسمى CHOOSE_TEST يقوم باختبار الحالة العاطفية للرياضيين قبل لعب البطولة لتحديد استقرارهم النفسي، سيتم الإشارة إلى المشاعر السلبية مثل الحزن والغضب على أنها حالة نفسية غير طبيعية، والمشاعر الإيجابية مثل مشاعر الفرح والطبيعي سيشار إليه بالحالة النفسية العادية. يقوم النظام بالكشف عن الحالة النفسية للاعب من خلال الصوت ومميزاته. قمنا ببناء نموذجين لهذا النظام: نموذج DL_CHOOSE المبني على الشبكات العصبية الالتفافية ونموذج ML_CHOOSE المبني على مفاهيم (ResNet50) المدربة مسبقًا ونقل التعلم. يتمتع النظام بمعدل دقة ٩٧٪ باستخدام الشبكات العصبية الالتفافية و٨٧٪ باستخدام ResNet50 لأغراض التمييز.

الكلمات المفتاحية: تمييز مشاعر الانسان، معاملات درجة النغم، التعلم العميق، ResNet50، الشبكات العصبية الالتفافية، تجميع البيانات

١.المقدمة

تحديد مشاعر الكلام هو عملية اكتشاف المشاعر البشرية من خلال إشارات الكلام. أصبحت هذه التقنية ذات أهمية متزايدة في نظام التفاعل بين الإنسان والحاسوب. يعد التعرف على عاطفة الكلام (Speech Emotion Recognition (SER) مهمة صعبة في مجال الذكاء الاصطناعي والتعلم الآلي. لقد طور الباحثون طرقًا مختلفة لتحديد العواطف من إشارات الكلام مع مرور الوقت [١]. العاطفة هي تجربة واعية قصيرة نسبيًا وتتميز بنشاط عقلي مكثف وإما بدرجة عالية من المتعة أو عدم الرضا، وهي حالة معقدة تؤدي إلى تغيرات جسدية ونفسية تؤثر على الفكر والسلوك. المشاعر هي حالات نفسية وفسيولوجية معقدة يمكن تحفيزها عن طريق محفزات داخلية أو خارجية. وهي تشمل مجموعة واسعة من المشاعر، بما في ذلك السعادة والحزن والغضب والخوف والمفاجأة. ويمكن نقل هذه المشاعر من خلال تعبيرات الوجه، ولغة الجسد، ونبرة الصوت، والسلوك [٢]. يتمتع نظام التعرف على مشاعر الكلام بكفاءة كافية لتمكين التواصل بين البشر والآلات. هذا النظام، المعروف باسم SER، قادر على تصنيف الحالات العاطفية مثل

الحزن والغضب والحياة الطبيعية والسعادة والمفاجأة بناءً على كلمات المتحدث. نظرًا لأن البشر يستخدمون اللغة الصوتية في المقام الأول للتواصل، فقد قام الباحثون باستكشاف طرق لتسهيل التواصل اللفظي مع الآلات. من خلال التعرف على الإدراك البشري من خلال SER، يمكن للآلات فهم البشر بشكل أفضل والاستجابة بشكل مناسب. تتيح هذه التقنية للآلات التعرف على المشاعر البشرية والتفاعل وفقًا لذلك [٣]. يعتمد نجاح نموذج التعلم الآلي بشكل كبير على ميزاته. لاستخراج المعلومات، يتم إجراء عملية استخراج الميزات لتعزيز التشابه والاختلاف بين الفئات، وتقليل أبعاد البيانات والحسابات. تلعب الميزات المستخرجة دورًا حاسمًا في التصنيف، ويهدف البحث إلى تحسين كفاءة التعرف على الكلام باستخدام طرق الاستخراج المقترحة [٤]. تحتوي الإشارات الصوتية على معلمات مختلفة يمكنها استنتاج الحالة العاطفية للمتحدث. يعد اختيار الميزة المناسبة أمرًا حيويًا في تحديد المشاعر. يتم تصنيف ميزات الكلام إلى أربع مجم[illegible] بناءً [illegible] الطاقة المالب (Teager Operator (TEO)



الشكل (١): أنواع ميزات إشارات الكلام

بعد استخراج الميزات، تتضمن الخطوة التالية تنفيذ خوارزمية تصنيف لتحديد المشاعر في الكلام بناءً على الإشارات الصوتية. ولتحقيق ذلك، استخدم الباحثون العديد من خوارزميات تصنيف التعلم الآلي بما في ذلك SVM وDT وRF، بالإضافة إلى خوارزميات التعلم العميق مثل CNN وLSTM. وكانت النتائج التي تم الحصول عليها مثيرة للإعجاب. بالإضافة إلى ذلك، تم دمج مصنفات مختلفة لتحسين دقة التعرف على المشاعر أو تمييزها في الكلام [٦][٧].

٢.النظام المقترح

تم اقتراح نموذج للذكاء المحوسب للتمييز بين المشاعر الطبيعية وغير الطبيعية لدى الإنسان. وقد تم تطوير النظام عبر عدة مراحل، بدءاً بإنشاء مجموعات البيانات، يليها استخلاص الخصائص، وأخيراً عملية التصنيف باستخدام التقنيات الذكية.

٢-١ تهيئة قاعدة البيانات

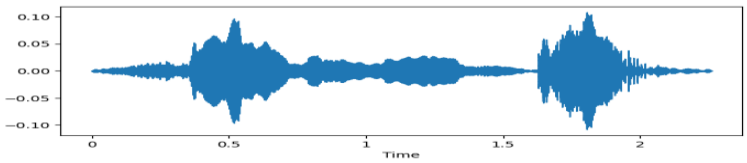
في نظام CHOOSE_TEST، تعتمد أهلية اللاعب للعب على مستوى هدوئه وحالته العاطفية. تتطلب أي رياضة التركيز والتوازن العاطفي ورباطة الجأش. لتحديد مدى هدوء اللاعب، يقوم المدرب بطرح أسئلة استفزازية على كل لاعب على حدة قبل ثلاث ساعات من المباراة خلال ما يسمى "فترة بداية الحمى"، ويقوم النظام بعد ذلك بتحديد اللاعبين الذين يظهرون الهدوء والتوازن ويرشحهم للعب في مرحلة حاسمة في البطولة. تتكرر هذه العملية خلال "فترة حمى البداية" لضمان اختيار اللاعبين الذين يظهرون الاستقرار العاطفي فقط للعب في هذه المرحلة الحاسمة من البطولة. في النظام المقترح، تم إنشاء مجموعة بيانات باللغة العربية الفصحى، تسمى "CHOOSE"، وتتكون من ٥٠٠ تسجيل صوتي يضم حالتين عاطفيتين (طبيعية وغير طبيعية) لخمسة لاعبين رياضيين (جميعهم ذكور) تتراوح أعمارهم بين ١٨ إلى ٢٣ عامًا. وتجرى بعض المعالجة الأولية على مجموعات البيانات التي تحتاج إلى معالجة أولية قبل إدخالها في مرحلة استخلاص المعلومات وتصنيفها.

٢-١-١ المعالجة الأولية

تعتبر المعالجة الأولية لبيانات الملفات الصوتية خطوة أساسية من أجل التعرف على عاطفة الكلام، لأنها تقوم بإعداد البيانات بشكل مناسب ومناسب للمرحلة اللاحقة، ومن أجل إعداد البيانات وإعدادها للتصنيف أو مرحلة التمييز، تمت المعالجة الأولية على مرحلتين:

المرحلة الأولى

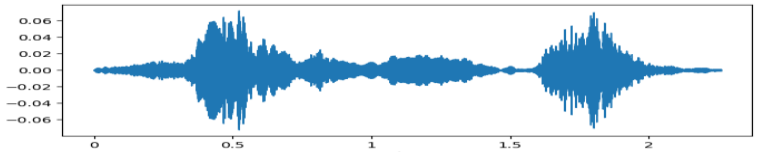
زيادة البيانات (Data augmentation) هي تقنية تستخدم لزيادة حجم مجموعة البيانات عن طريق إنشاء عينات جديدة من العينات الموجودة. في مجموعات بيانات المشاعر اللفظية، يمكن استخدام زيادة البيانات لتحسين أداء نماذج التعرف على المشاعر عن طريق إدخال اختلافات في إشارات الكلام [٩]، ويمثل الشكل (٢) إشارة صوتية قبل إجراء زيادة البيانات.



الشكل (٢) الإشارة الأصلية قبل عملية زيادة البيانات

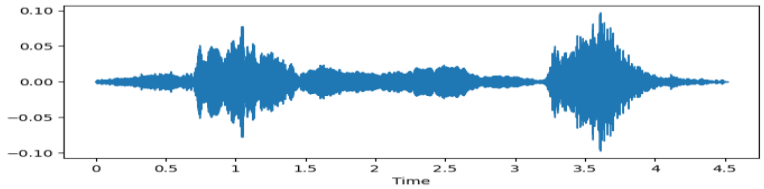
ومن التقنيات المستخدمة لزيادة بيانات الملفات الصوتية:

١- تغيير طبقة الصوت: ويتضمن تغيير طبقة إشارة الكلام مع الحفاظ على مدتها ومضمونها. يمكن التلاعب بتبديل طبقة الصوت لمحاكاة التغيرات في الشدة العاطفية أو لإنشاء عينات جديدة بنطاقات مختلفة من طبقات الصوت لها تأثير يشبه الصدى [١٠]. كما هو مبين في الشكل (٣).



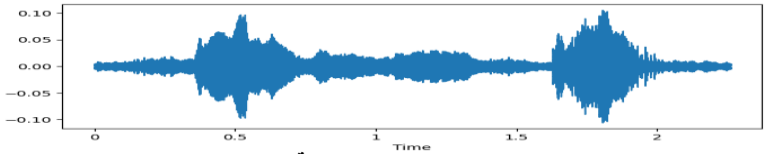
الشكل (٣): إشارة ضارية

٢. تمديد الوقت: يتضمن تغيير سرعة إشارة الكلام مع الحفاظ على طبقه الصوت ومضمونه. يمكن استخدام تمدد الوقت لمحاكاة التغيرات في معدل التحدث أو لإنشاء عينات جديدة بمدد مختلفة [٩]. كما هو مبين في الشكل (٤).



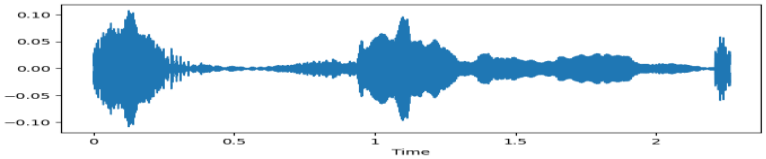
الشكل (٤): إشارة ممتدة

٣. إضافة الضوضاء: يتضمن ذلك إضافة الضوضاء البيضاء إلى إشارة الكلام لمحاكاة بيئات العالم الحقيقي أو لإنشاء عينات جديدة بمستويات مختلفة من الضوضاء في الخلفية [١٠]. كما هو مبين في الشكل (٥).



الشكل (٥): إشارة مشوشة

٤. تحويل الإشارة الصوتية: يتضمن ذلك تغيير الطيف الترددي لإشارة الكلام مع الحفاظ على بنيتها الزمنية. ويمكن استخدام الالتواء الطيفي لمحاكاة التغيرات في شكل المجرى الصوتي أو لتوليد عينات جديدة ذات خصائص طيفية مختلفة [٩]، كما هو موضح في الشكل (٦).

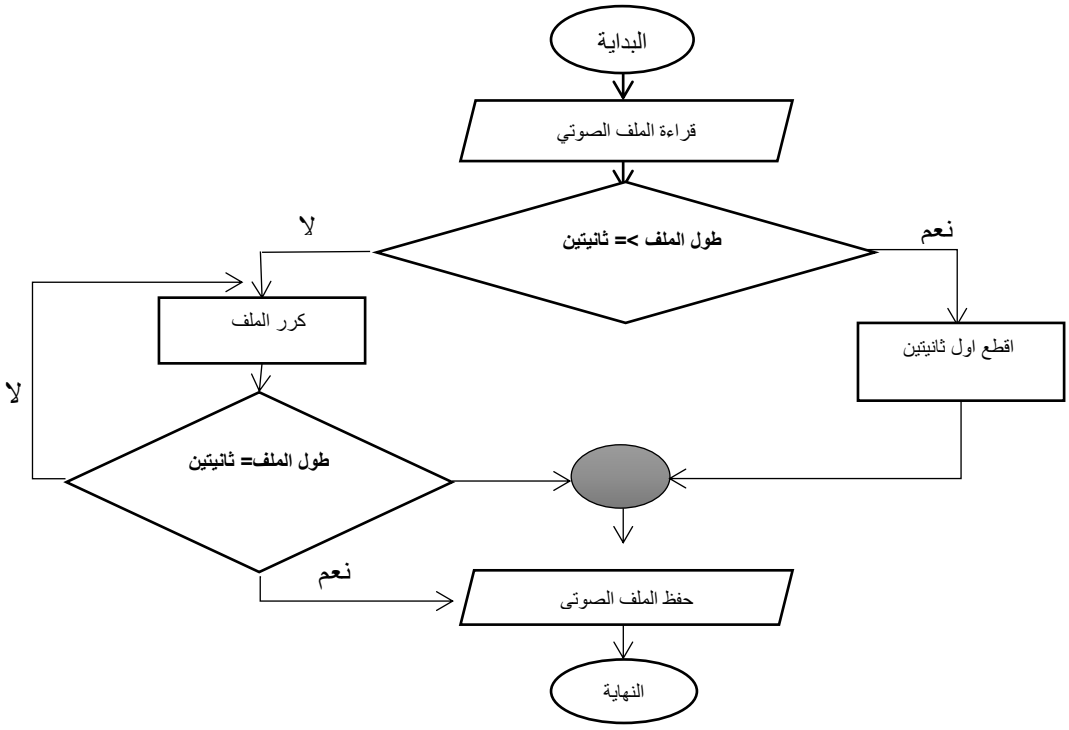


الشكل (٦): إشارة متحولة

في نظام CHOOSE_TEST المقترح، أنشأنا مجموعة البيانات الخاصة بنا واستخدمنا أربع تقنيات مختلفة لزيادة البيانات لزيادة حجم مجموعة البيانات وتجنب التجهيز الزائد (overfitting).

المرحلة الثانية

توحيد طول الملف ومعدل العينة حيث تتضمن هذه العملية ضبط معدل أخذ العينات للملفات الصوتية ليكون (٢٢٠٥٠) عينة في الثانية لجميع الملفات، وكذلك توحيد أطوال الملفات الصوتية للحصول على أحجام متساوية للعينات الصوتية وبالتالي الحصول على النتائج المطلوبة. تم توحيد الملفات الصوتية وتحديد الطول المطلوب للملف الصوتي ب(ثانيتين). إذا كان طول الملف الصوتي أكثر من ثانيتين، فسيتم خصم أول ثانيتين فقط. أما إذا كان طول الملف أقل من ثانيتين، فسيتم تكرار الملف الصوتي حتى يصل إلى الطول المطلوب. تظهر عملية توحيد طول الملفات في الشكل (٧).



الشكل (٧): مخطط انسيابي لعملية توحيد طول الملف

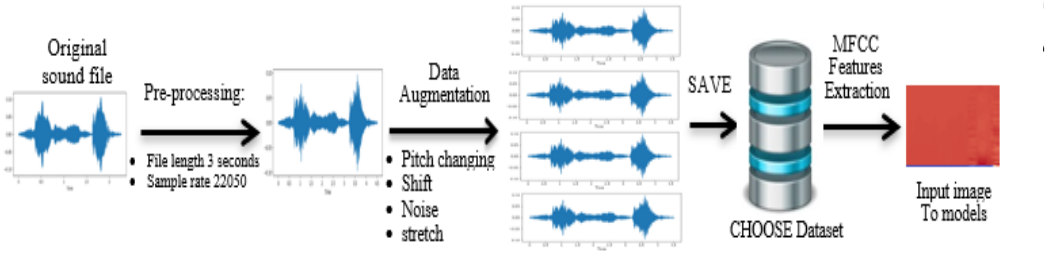
يوضح الجدول (١) مجموعة البيانات المستخدمة في نظامنا المقترح والحالات العاطفية في كل مجموعة

جدول (١) قاعدة البيانات التي سميت (CHOOSE)

العينة	طبيعي	غير طبيعي
الاصلية	٥٠	٥٠
ضارية	٥٠	٥٠
ممتدة	٥٠	٥٠
مشوشة	٥٠	٥٠
متحولة	٥٠	٥٠
المجموع	٢٥٠	٢٥٠

٢-٢ استخلاص الميزات باستخدام معاملات درجة النغم (MFCCs)

تعد معاملات درجة النغمة، تمثيلًا مضغوطًا لطيف الإشارة الصوتية الذي يأخذ في الاعتبار الإدراك البشري غير الخطي لطبقة الصوت. الخطوة الأولى في أي نظام للتعرف على عواطف الكلام هي استخراج الميزات عن طريق تحديد مكونات الإشارة الصوتية التي تعتبر جيدة لتحديد المحتوى العاطفي وإزالة جميع المعلومات الأخرى مثل ضوضاء الخلفية واللغة [٨]. يتم تصفية الأصوات التي يصدرها الإنسان من خلال شكل المجرى الصوتي الذي يشمل اللسان والأسنان، مما يحدد شكل الصوت الذي يخرج. إن تحديد هذا الشكل بدقة يعطي تمثيلاً دقيقًا للصوت الذي يتم إنتاجه. يتجلى غلاف طيف الطاقة قصير الأمد في شكل القناة الصوتية، وتمثل معاملات درجة النغم هذا الظرف بدقة. تم استخدام هذه المعاملات على نطاق واسع في التعرف التلقائي على عواطف الكلام منذ أن قدمها ديفيس وميرميلستين في الثمانينات [١١].



الشكل (٨): استخلاص الميزات من قاعدة البيانات (CHOOSE) باستخدام معاملات درجة النغم

٢-٣ النظام المقترح

اولاً: نظام DL_CHOOSE المعتمد على الشبكات العصبية الالتفافية

يُعرف نموذج الشبكات العصبية الالتفافية المصمم للتمييز بين مشاعر اللاعب بنموذج DL_CHOOSE.وهو نموذج متسلسل تم إنشاؤه باستخدام مكتبة (Keras) في لغة بايثون، وهيكله جدول (٢). ويتكون النموذج من مرحلتين: مرحلة استخلاص الميزات ومرحلة التصنيف.

جدول (٢) هيكلية نموذج CNN_CHESS

Layer	type	details
1	Conv2d	6 filters + ReLU + Stride=1 + kernel width=5
2	Conv2d	16 filters + ReLU + Stride=1 + kernel width=5
3	Conv2d	32 filters + ReLU + Stride=1 + kernel width =5
4	Conv2d	64 filters + ReLU + Stride=1 + kernel width =3
	MaxPooling	Global 2×2
5	Dense	128 units +ReLU activation
6	Softmax	2 classes classification

مرحلة الاستخراج تتضمن (٤) طبقات بأبعاد (٦٤*٦٤*٣). الطبقة الأولى من نوع (Conv2D) وتتكون من (٦) مرشحات. تم استخدام مرشح (Kernel) لعملية الالتفاف، ممثلاً بمصفوفة ثنائية بحجم ٥*٥، وتم تطبيق دالة التنشيط من نوع ReLU بالإضافة إلى ذلك، تم استخدام مصفوفة (MaxPooling2D) بحجم ٢*٢ لتبسيط المعلومات التي تم الحصول عليها بواسطة الطبقة الالتفافية. أما الطبقة الثانية فهي أيضاً

من النوع (Conv2D) وتتكون من (١٦) مرشحاً. تم استخدام مرشح (Kernel) للالتفاف بحجم مصفوفة ثنائية ٥*٥، وتم تطبيق وظيفة التنشيط من نوع ReLU تم استخدام مصفوفة (MaxPooling2D) مرة أخرى بحجم ٢*٢. أما الطبقة الثالثة فهي أيضاً من النوع Conv2D وتتكون من (٣٢) مرشحاً. تم استخدام مرشح (Kernel) للالتفاف بحجم مصفوفة ثنائية ٥*٥، وتم تطبيق وظيفة التنشيط من نوع ReLU.ومرة أخرى تم استخدام مصفوفة (MaxPooling2D) بحجم ٢*٢. والطبقة الرابعة هي أيضاً من النوع الالتفافي ثنائي الأبعاد (Conv2D) وتتكون من (٦٤) مرشحاً. تم استخدام مرشح (Kernel) للالتفاف بحجم مصفوفة ثنائية ٣*٣، وتم تطبيق وظيفة التنشيط من نوع ReLU وأخيراً تم استخدام مصفوفة (MaxPooling2D) أخرى بحجم ٢*٢ لتبسيط المعلومات التي حصلت عليها هذه الطبقة أيضاً.

وتتكون مرحلة التعرف على فئة العاطفة من (٣) طبقات. الطبقة الأولى من النوع (Flatten)، أما الطبقة الثانية فهي من النوع التقديمي المتصل بالكامل (Dense) وتحتوي على ١٢٨ عقدة. يتم استخدام وظيفة التنشيط من نوع ReLU، مع تحديد قيمة تسرب بمقدار ٠,٥. تحتوي طبقة الإخراج على خمس عقد تمثل الحالات العاطفية الخمس التي يجب تمييزها. يتم استخدام وظيفة تنشيط (SoftMax) في هذه الطبقة.

ثانيا: نموذج ML_CHOOSE يعتمد على نماذج الشبكات العصبية الالتفافية المدربة مسبقًا

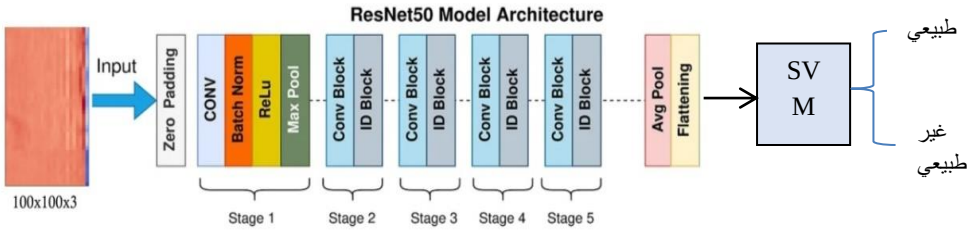
نموذج ML_CHOOSE هو نموذج تعلم النقل تم تدريبه مسبقًا ويستخدم للتعرف على عواطف الكلام. لقد تم تدريبه على قاعدة بيانات صور كبيرة تسمى (ImageNet) ويستخدم ResNet50 للتعرف على فئتين في مجموعة بيانات (CHOOSE) الخاصة بالكلام. تتطلب طبقة الإدخال معلومات حول نموذج الإدخال، بينما تعمل الطبقات الأخرى بناءً على القيم والأوزان التي تم تدريبها مسبقًا. خطوة استخراج الميزات لنموذج ML_CHOOSE المبنية على ResNet50 تتضمن خمس مراحل كما هو موضح في الجدول (٣). تُستخدم الطبقات المخفية لاستخراج الميزات من الصورة الطيفية المدخلة باستخدام معاملات درجة النغم ودمجها مع مصنفات التعلم الآلي (SVM، RF، DT) لتصنيف المدخلات إلى واحدة من فئتين نفسيتين عاطفيتين: عادية وغير طبيعية. تشير الحالة الطبيعية إلى المشاعر الطبيعية، بينما تشير الحالة غير الطبيعية إلى المشاعر غير المستقرة مثل الغضب أو الحزن.

جدول (٣) هيكلية نموذج TL_CHOOSE

Stages	Details
first	- Input layer: 100x100x3 - Convolutional layer (64 filters, kernel size 7x7): 50x50x64 - Max pooling layer (kernel size 3x3): 24x24x64
second	- Residual block (64 filters): 24x24x64 - Residual block (64 filters): 24x24x64 - Residual block (64 filters): 24x24x64

third	- Convolutional layer (128 filters, stride=2, kernel size 1x1): 12x12x128 - Residual block (128 filters): 12x12x128 - Residual block (128 filters): 12x12x128 - Residual block (128 filters): 12x12x128
fourth	- Convolutional layer (256 filters, stride=2, kernel size 1x1):6X6X256 - Residual block(256filters) :6X6X256 - Residual block(256filters) :6X6X256 - Residual block(256filters) :6X6X256
fifth	Convolutional layer (512filters, stride=2, kernel_size=1*1):3X3X512. - Residual block(512filters):3X3X512. - Residual block(512filters):3X3X512. - Average pooling layer (kernel size=7×7):1×1×2048. - **Fully connected layer:1000 classes. (This layer will be removed)**

يحتوي ResNet-50 على ٥٠ طبقة ويتضمن الاتصالات المتبقية بين الطبقات. تساعد هذه الروابط في تقليل الخسارة، والحفاظ على اكتساب المعرفة، وتحسين الأداء أثناء التدريب. يشير الاتصال المتبقي داخل الطبقة إلى أن مخرجات الطبقة عبارة عن التفافات لمدخلاتها بالإضافة إلى مدخلاتها [١٢]. يعتمد نموذج SERTL على ResNet50 ويستخدم صورة طيفية تم استخراجها بواسطة معاملات درجة النغم وبحجم (١٠٠*١٠٠*٣) كمدخل، كما هو موضح في الشكل (٩).



الشكل (٩): نموذج TL_CHOOSE الذي يعتمد على هيكلية ResNet50

٣.النتائج

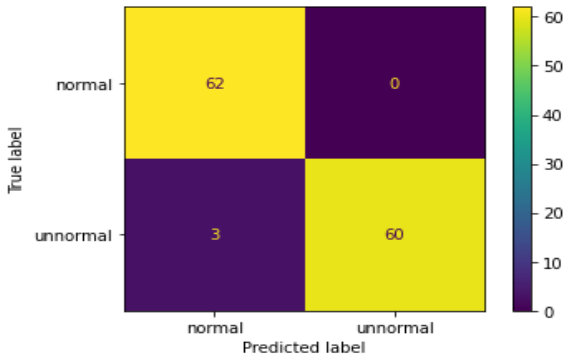
أولا: نتائج نموذج DL_CHOOSE

في هذا النموذج تم استخدام خوارزمية الشبكة العصبية الالتفافية للتعرف على حالتين نفسيتين (طبيعية، غير طبيعية) في استخدام قاعدة البيانات CHOOSE. تم تدريب شبكة الشبكات العصبية الالتفافية على مجموعة بيانات CHOOSE مع معاملات درجة النغم وعددها (٢٦)، باستخدام حجم دفعة (٣٢) وعصر (١٥٠). باستخدام مقياس التباين (درجة F1، الدقة، الاستدعاء)، حقق نموذج DL_CHOOSE دقة التصنيف ٩٧ % كما هو مبين في الجدول (٤).

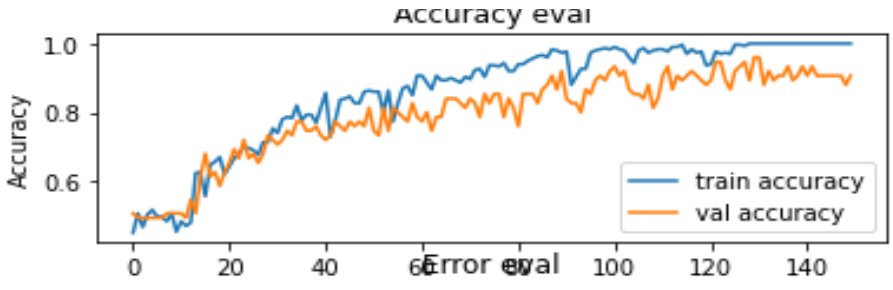
جدول (٤) مقاييس دقة نموذج DL_CHOOSE في مجموعة بيانات "CHOOSE".

psychological states	precision	recall	f1-score	support
Normal	0.94	1.00	0.96	63
Abnormal	1.00	0.96	0.97	62
ACCURACY				97%

تعرض مصفوفة الارتباك في الشكل (١٠) القيم الفعلية والقيم المتوقعة الناتجة عن مصنف نموذج الشبكة العصبية الالتفافية (DL_CHOOSE). في الشكل (١١) انظر منحنى التعلم الذي يوضح دقة التدريب والتحقق لنموذج DL_CHOOSE، بعد (١٥٠) دورة تدريبية بحجم (٣٢) دفعة.



الشكل (١٠): مصفوفة الارتباك لنموذج DL_CHOOSE



الشكل (١١): منحنى التعلم والدقة لنموذج DL_CHOOSE

ثانيا: نتائج نموذج ML_CHOOSE

يستخدم نموذج ML_CHOOSE أسلوب نقل التعلم من خلال استخدام شبكة ResNet50. ويتم الحصول على النتائج حسب تصنيف التباين كما هو موضح في الجدول (٥) والجدول (٦).

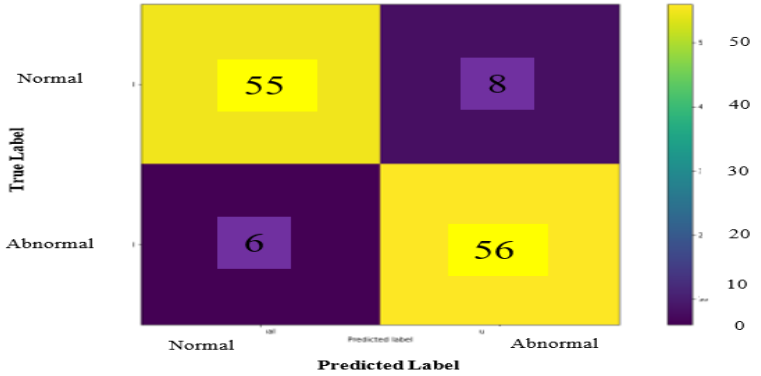
الجدول (٥): مقاييس دقة نموذج ML_CHOOSE في مجموعة بيانات "CHOOSE" باستخدام مصنف RF

psychological states	precision	recall	f1-score	support
Normal	0.90	0.85	0.88	٦٣
Abnormal	0.85	0.90	0.87	٦٢
ACCURACY			87%	١٢٥

الجدول (٦): مصنفات نموذج ML_CHOOSE: تقييم الدقة باستخدام مصنفات مختلفة

Classifier	Accuracy
RF	**87%**
SVM	**86%**
RF	**75%**

حقق نموذج مصفوفة الارتباك ML_CHOOSE الذي استخدم شبكة ResNet50 دقة تصنيف (٨٧٪) بواسطة المصنف Random Forest (RF) لـ (٢) حالات عاطفية نفسية موضحة في الشكل (١٢).



الشكل (١٢): مصفوفة الارتباك لنموذج ML_CHOOSE المبنى على ResNet50

الاستنتاج

في هذا البحث، قمنا بتطوير نموذجين للتعلم الآلي لتمييز الحالات العاطفية والنفسية للاعبي فريق الشطرنج بجامعة نينوى، DL_CHOOSE باستخدام التعلم العميق، وML_CHOOSE، نموذج التعلم الآلي باستخدام ResNet50. استخدمنا مجموعة بيانات CHOOSE. تمت معالجة الملفات الصوتية وتحويلها إلى صورة ميزات باستخدام معاملات درجة النغم (MFCC) بعد إجراء المعالجة المسبقة وزيادة البيانات. حققت DL_CHOOSE دقة بلغت ٩٧٪، وانخفضت الدقة إلى ٨٧٪ في ML_CHOOSE المعتمدة على شبكة ResNet50. استخدمنا ثلاث مصنفات في نموذج ML_CHOOSE وهي (SVM,DT,RF)، ووجدنا أن مصنفات RF وSVM أنتجت نتائج إيجابية عند استخدامها كمصنفات لنظام التعرف على عواطف الكلام.

المصادر

[1] L. K. Kuhn, “Emotion recognition in the human face and voice.” Brunel University London, 2015.

[2] M. Cabanac, “What is emotion?,” *Behav. Processes*, vol. 60, no. 2, pp. 69–83, 2002.

[3] H. S. Kumbhar and S. U. Bhandari, “Speech emotion recognition using MFCC features and LSTM network,” in *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, 2019, pp. 1–3.

[4] J.-H. Yeh, T.-L. Pao, C.-Y. Lin, Y.-W. Tsai, and Y.-T. Chen, “Segment-based emotion recognition from continuous Mandarin Chinese speech,” *Comput. Human Behav.*, vol. 27, no. 5, pp. 1545–1552, 2011.

[5] S. Wu, T. H. Falk, and W.-Y. Chan, “Automatic speech emotion recognition using modulation spectral features,” *Speech Commun.*, vol. 53, no. 5, pp. 768–785, 2011.

[6] S. T. Alam Monisha and S. Sultana, “A Review of the Advancement in Speech Emotion Recognition for Indo-Aryan and Dravidian Languages,” *Adv. Human-Computer Interact.*, vol. 2022, 2022.

[7] M. Valstar *et al.*, “Avec 2016: Depression, mood, and emotion recognition workshop and challenge,” in *Proceedings of the 6th international workshop on audio/visual emotion challenge*, 2016, pp. 3–10.

[8] S. B. Jagtap, K. R. Desai, and M. J. K. Patil, “A Survey on Speech Emotion Recognition Using MFCC and Different classifier,” in *8th national conference on emerging trends in engg and technology*, 2018.

[9] B. T. Atmaja and A. Sasou, “Effects of Data Augmentations on Speech Emotion Recognition,” *Sensors*, vol. 22, no. 16, pp. 1–13, 2022, doi: 10.3390/s22165941.

[10] H. Alsayadi, A. Abdelhamid, I. Hegazy, and Z. Taha, “Data Augmentation for Arabic Speech Recognition Based on End-to-End Deep Learning,” *Int. J. Intell. Comput. Inf. Sci.*, vol. 21, no. 2, pp. 50–64, 2021, doi: 10.21608/ijicis.2021.73581.1086.

[11] E. Vázquez-Cano, E. López-Meneses, and M. . L. Sevillano García, “La competencia digital y las diferencias de género entre los estudiantes universitarios,” *Tecnol. Innovación e Investig. en los procesos enseñanza-aprendizaje*, no. 4, pp. 1929–1936, 2016.

[12] L. Ali, F. Alnajjar, H. Al Jassmi, M. Gocho, W. Khan, and M. A. Serhani, “Performance Evaluation of Deep CNN-Based Crack Detection and Localization Techniques for Concrete Structures,” 2021.

[13] M. El Ayadi, M. S. Kamel, and F. Karray, “Survey on speech emotion recognition: Features, classification schemes, and databases,” *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.

Selecting Sports Players Using Intelligent Techniques

Assistant Lecturer Mohammed N. Kannah

mohammed_naif85@nan.epedu.gov.iq

Abstract: Speech is one of the most natural forms of expression. As technology advances, the interaction between humans and machines is becoming more sophisticated. Research has shown that emotions can be detected not only through facial expressions but also through sounds. Like

facial features, auditory characteristics such as pitch and intensity can provide valuable information about a person's emotional state. Studies suggest that it is possible to build intelligent systems that recognize individuals' emotional states through their voices, distinguishing between anger, sadness, happiness, fear, normal state, and others. These systems can be used in various applications such as driver safety systems and customer service systems. In this study, we aim to apply emotion recognition systems in the sports field. The developed system called CHOOSE_TEST that tests the emotional state of CHOOSE players before they play a tournament to determine their psychological stability, Negative feelings such as sadness and anger will be referred to as an ABNORMAL psychological state, and positive feelings such as feelings of joy and natural will be referred to as NORMAL psychological state. The system detects the player's psychological state through sound and its features. We built two models for this system: DL_CHOOSE model based on CNN and ML_CHOOSE model based on pre-trained (ResNet50) and transfer learning concepts. The system has an accuracy rate of 97% using CNN and 87% using ResNet50 for discrimination purposes.

Keywords: SER, MFCC, Deep Learning, ResNet50, CNN, Data Augmentation