



Constitution des bases de données

Modules de la base RAVIOLI

- École : micro
- Repas : micro
- Itinéraire : micro discret
- 24H : micro-cravate

1	Nom du fichier	Locuteur	Injonctive?	Categorie	Start	End	Durée	Texte
2	24H_apresmidtravail_4.wav	NR390	oui	excuse	189.69	191.75	2.06	pardon je te le mets déjà sur les pieds je suis désolée
3	24H_apresmidtravail_4.wav	NR390	oui	apostrophe	346.9	347.33	0.43	Sarah
4	24H_apresmidtravail_4.wav	NR390	oui	excuse	390.97	391.86	0.89	pardon excusez -moi
5	24H_apresmidtravail_4.wav	NR390	oui	refus	498.53	500.93	2.4	bah vu comment il est cintré excuse -moi
6	24H_apresmidtravail_4.wav	collègue 4	oui	stop	551.36	552.15	0.79	attends attends attends attends
7	24H_apresmidtravail_4.wav	NR390	oui	UNDEF	574.24	574.45	0.21	non
8	24H_apresmidtravail_4.wav	NR390	oui	UNDEF	574.45	574.88	0.43	t' inquiète
9	24H_apresmidtravail_4.wav	collègue 4	oui	UNDEF	602.63	604.55	1.92	oui mais faut les serrer Pamela

Figure: Tableur des audio injonctifs

Extraction de données

Hauteur (PI)

- logiciel Praat
- calculer sur des trames de $10ms$
- la vitesse (A) et l'accélération (D) sont calculés avec HTK

Énergie logarithmique (E)

- Logiciel HTK
- calculer toutes les $10ms$ sur des fenêtres d'analyse de Hamming de $30ms$
- La vitesse (A) et l'accélération (D) sont calculés avec HTK

Méthodes SVM et KNN appliqués sur la première base

Dans le cas de la validation croisée (2-folds)

Configs	E	E,E _D	E,E _{DA}	PI	E,E _D ,PI _D	PI,E,E _{DA} ,PI _{DA}
N	3	5	7	27	123	11
TC (%)	[50.75, 69.89]	[50.75, 62.75]	[53.76, 61.73]	[49.74, 65.81]	[49.24, 68.36]	[49.50, 65.81]

Table: Résultats obtenus avec la méthode KNN en utilisant 2-fold ([1-fold,2-fold])

Configs	E	E,E _D	E,E _{DA}	PI	E,E _D ,PI _D	PI,E,E _{DA} ,PI _{DA}
C,γ	0.01,0.01	0.01,0.01	0.1,0.01	10,0.01	0.01,0.01	0.5,0.01
TC (%)	[49.74, 71.93]	[49.74, 64.28]	[49.74, 61.73]	[34.18, 50.25]	[33.36, 51.25]	[51.25, 71.93]

Table: Résultats obtenus avec la méthode SVM en utilisant 2-fold([1-fold,2-fold])

Méthodes SVM et KNN appliqués sur la première base

	INJ	NINJ
1-fold (minutes)	1.2	0.909
2-fold (minutes)	1.3215	3.463

Table: La durée en minute des audio dans chaque fold

Configs	Méthode KNN		Méthode SVM	
	groupe1	groupe2	groupe1	groupe2
$E, E_{DA}, P_I, P_{I_{DA}}$	$\begin{pmatrix} 0 & 100 \\ 1 & 98 \end{pmatrix}$	$\begin{pmatrix} 44 & 53 \\ 14 & 85 \end{pmatrix}$	$\begin{pmatrix} 1 & 99 \\ 0 & 99 \end{pmatrix}$	$\begin{pmatrix} 54 & 43 \\ 11 & 88 \end{pmatrix}$
E	$\begin{pmatrix} 2 & 98 \\ 0 & 99 \end{pmatrix}$	$\begin{pmatrix} 31 & 66 \\ 7 & 92 \end{pmatrix}$	$\begin{pmatrix} 1 & 99 \\ 0 & 99 \end{pmatrix}$	$\begin{pmatrix} 25 & 72 \\ 1 & 98 \end{pmatrix}$
P_I	$\begin{pmatrix} 0 & 100 \\ 0 & 99 \end{pmatrix}$	$\begin{pmatrix} 43 & 54 \\ 13 & 86 \end{pmatrix}$	$\begin{pmatrix} 13 & 87 \\ 15 & 84 \end{pmatrix}$	$\begin{pmatrix} 26 & 71 \\ 15 & 84 \end{pmatrix}$
P_I, E_D, P_{I_D}	$\begin{pmatrix} 0 & 100 \\ 1 & 98 \end{pmatrix}$	$\begin{pmatrix} 39 & 58 \\ 4 & 95 \end{pmatrix}$	$\begin{pmatrix} 4 & 96 \\ 1 & 98 \end{pmatrix}$	$\begin{pmatrix} 13 & 84 \\ 46 & 53 \end{pmatrix}$

Table: Les matrices de confusion obtenues par les méthodes KNN et SVM dans le cas de 2-folds appliquées sur la première base

Méthode KNN appliqué sur la deuxième base

Configs	2-folds			5-folds		
	N	TC(%)	moyen TC(%)	N	TC(%)	moyen TC(%)
E,PI,E _{DA} ,PI _{DA}	21	[65.82, 61.20]	63.51	7	[63, 64, 65.50, 62, 64]	64.17
E	95	[55.81, 56.48]	56.15	21	[51.37, 56.04, 55.21, 59.34, 58.79]	56.15
E, E _D	49	[58.79, 60.43]	59.61	77	[55.76, 59.89, 59.06, 56.31, 65.93]	59.39
E, E _A	45	[58.79, 62.96]	60.87	91	[58.51, 62.08, 58.79, 62.08, 63.73]	61.04
E,E _{DA}	69	[53.07, 55.93]	54.50	77	[59.06, 63.73, 59.89, 59.89, 64.83]	61.48
PI	9	[60.95, 55.71]	58.35	9	[56.59, 60.98, 56.86, 56.04, 56.59]	57.41
PI, PI _D	33	[61.64, 56.70]	59.175	43	[54.67, 62.36, 57.96, 57.14, 59.89]	58.40
PI, PI _A	55	[60.54, 56.48]	59.17	25	[54.94, 62.08, 61.81, 57.41, 53.84]	58.02
PI, E	63	[54.84, 54.84]	54.84	13	[56.31, 62.91, 60.71, 58.24, 54.12]	58.46
PI, PI _D , E _D	85	[62.19, 58.13]	60.16	9	[56.86, 59.89, 57.69, 58.24, 59.34]	58.40

Table: Résultats obtenus avec la méthode KNN en utilisant 2-folds et 5-folds ([1fold,...,5fold]) appliqués sur la base complète

Methode SVM appliquée sur la dixième base

Configs	2-folds				5-folds			
	C	γ	TC	moyen TC	C	γ	TC	moyen TC
E,PI,E _{DA} ,PI _{DA}	1	1	[58.58, 63.49]	61.04	0.1	10	[62, 63.5 64.5, 63, 51]	63
E	1	1	[56.15, 56.59]	56.37	10	10	[51.64, 58.79, 54.67, 57.69, 60.71]	56.70
E, E _D	10	10	[55.91, 59.34]	58.62	10	10	[53.57, 59.06, 56.31, 60.71, 62.91]	58.51
E, E _A	10	10	[56.92, 58.35]	57.63	10	10	[53.29, 58.24, 56.04, 57.41, 61.81]	57.36
E,E _{DA}	1	10	[57.25, 59.23]	58.24	1	10	[55.49, 59.34, 57.40, 58.24, 61.53]	58.40
PI	10	10	[57.58, 52.74]	55.16	0.1	10	[52.47, 57.69, 56.31, 53.29, 54.67]	54.89
PI, PI _D	10	1	[59.67, 55.93]	57.80	1	1	[56.04, 60.71, 59.06, 56.86, 58.24]	58.18
PI, PI _A	10	1	[60.30, 58.02]	59.17	1	10	[55.49, 59.89, 59.34, 56.31, 54.94]	57.19
PI, E	1	10	[60, 57.80]	58.90	1	1	[53.02, 59.34, 55.49, 55.76, 53.57]	55.43
E, PI _D , E _D	0.1	10	[60, 57.80]	58.90	1	1	[58.79, 59.89, 58.51, 59.06, 60.16]	59.28

Table: Résultats obtenus avec la méthode SVM en utilisant 2-folds et 5-folds ([1fold,...,5fold]) appliqués sur la base complète

Méthodes KNN et SVM appliquées sur la base école

Configs	Méthode KNN			Méthode SVM			
	N	TC	Moyen TC	C	γ	TC	MoyenTC
E,PI,E _{DA} ,PI _{DA}	35	[61.35, 63.07]	62.21	10	0.01	[61.60, 62]	61.80
E	43	[56.60, 55.19]	55.90	0.1	1	[55.89, 57.88]	56.88
E,E _D	29	[57.67, 62]	59.84	1	10	[57.14, 59.49]	58.32
E,E _A	45	[58.92, 56.98]	57.95	10	10	[56.25, 58.42]	57.33
E,E _{DA}	151	[58.75, 63.62]	61.18	10	10	[57.14, 59.85]	58.49
PI	3	[55.71, 60.03]	57.87	0.1	0.1	[56.07, 57.52]	56.79
PI, PI _A	31	[56.78, 61.64]	59.21	10	0.1	[55.89, 60.57]	58.23
PI, PI _D	29	[57.85, 60.21]	59.03	1	0.1	[55.71, 58.06]	57.51
PI, PI _{DA}	49	[56.60, 61.11]	58.85	1	0.1	[55.89, 60.75]	58.32
PI, E	15	[56.96, 60.93]	58.94	1	0.1	[56.07, 56.80]	56.44
E, E _D , PI _D	201	[56.96, 62.36]	59.66	0.1	10	[59.99, 57.80]	58.90
E, E _D , PI _{DA}	201	[57.32, 64.33]	60.82	1	10	[56.42, 58.24]	57.33
E, E _{DA} , PI _{DA}	21	[60.71, 62.90]	61.80	1	10	[56.60, 57.88]	57.24

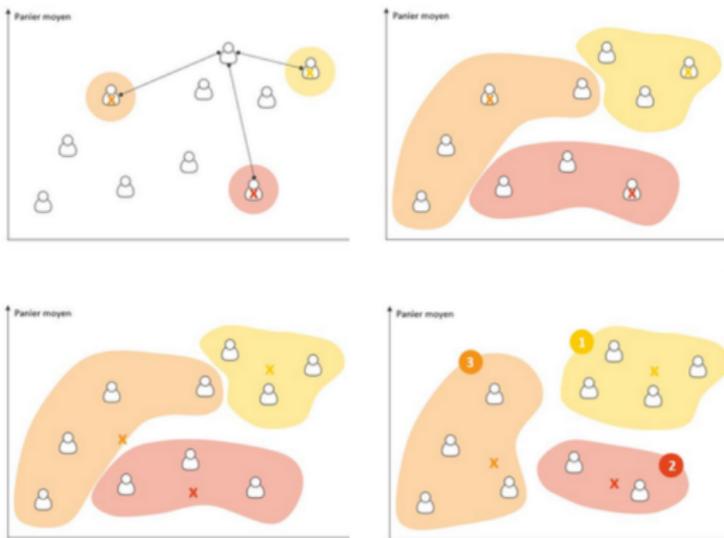
Table: Résultats obtenus avec la base **école** par les méthodes KNN et SVM en utilisant 2-fold

Méthodes KNN et SVM appliquées sur la base repas

Configs	Méthode KNN			Méthode SVM			
	N	TC	moyen TC	C	γ	TC	TCmoyen
E,PI,E _{DA} ,PI _{DA}	15	[61.65, 64.41]	63.03	0.1	0.1	[60.20, 64.88]	62.54
E	201	[50, 53.37]	51.68	0.1	0.1	[51, 52]	51.50
E,E _D	23	[55.21, 53.37]	54.29	1	10	[50.92, 54.60]	52.76
E,E _A	91	[52.14, 54.90]	53.52	1	0.1	[50.50, 51.50]	51
E,E _{DA}	251	[57.05, 57.97]	57.51	1	10	[50.92, 54.60]	52.76
PI	7	[57.97, 65.44]	57.20	0.1	0.01	[59.5, 60.12]	59.81
PI,PI _A	47	[58.89, 60.73]	59.81	10	1	[58.58, 63.49]	61.04
PI,PI _D	79	[58.58, 61.04]	59.81	0.1	0.01	[59.81, 62.57]	61.19
PI,PI _{DA}	29	[60.42, 63.49]	61.96	0.1	0.01	[59.81, 64.11]	61.96
PI, E	5	[58.89, 57.36]	58.12	0.1	0.01	[59.50, 60.12]	59.81
E, E _D ,PI _D	25	[57.05, 57.97]	57.51	1	1	[56.74, 57.36]	57.05
E, E _D ,PI _{DA}	51	[56.74, 61.34]	59.04	0.1	1	[56.74,62.26]	59.50
E, E _{DA} ,PI _{DA}	11	[58.89, 61.96]	62.07	0.1	1	[56.75,62.28]	59.51

Table: Résultats obtenus avec la base **repas** par les méthodes KNN et SVM en utilisant 2-fold

k-means



Méthode K-means

- Méthode d'apprentissage non supervisé
- Devise les données en k clusters cohérents

Figure: Principe de la méthode K-means [7]

Méthode k-means appliquée sur la deuxième base

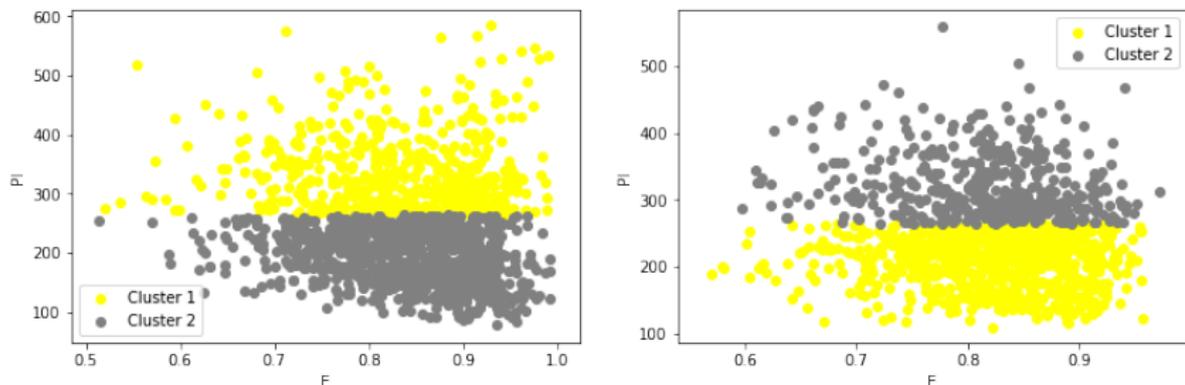


Figure: Répartition des valeurs injonctives (à gauche) et non injonctives (à droite) de E et PI avec 2-means

Méthode k-means appliquée sur la deuxième base

Configs	Méthode KNN			Méthode SVM			
	N	TC	moyen TC	C	γ	TC	TCmoyen
E,PI,E _{DA} ,PI _{DA}	13	[67.10, 71]	69.05	10	10	[67.57, 71.07]	68.03
E	17	[74, 74.50]	74.25	0.1	10	[71.25, 70.07]	70.66
E,E _D	15	[73.62, 70.86]	72.24	0.1	10	[70.86, 70.07]	70.47
E,E _A	7	[75.39, 71.85]	73.62	0.1	0.1	[71.45, 68.89]	70.17
E,E _{DA}	11	[74, 71]	72.50	0.1	10	[65, 71.07]	68.03
PI	3	[57, 55]	56	10	10	[65.21, 63.81]	67.75
PI,PI _A	21	[65.57, 66.66]	66.12	10	1	[64.49, 66.12]	65.30
PI,PI _D	93	[67.02, 65.45]	66.24	1	10	[64.49, 65.21]	64.85
PI,PI _{DA}	31	[66, 67]	66.50	10	1	[64.67, 66.30]	65.48
PI, E	23	[65, 67]	65.50	0.1	1	[64.67, 64.67]	64.67
E, E _D ,PI _D	59	[61.89, 60.87]	61.38	1	1	[56.74, 57.36]	57.05
E, E _D ,PI _{DA}	87	[62.83, 60.95]	61.89	0.1	1	[56.74,62.26]	59.50
E, E _{DA} ,PI _{DA}	77	[63.93, 64.49]	64.21	0.1	1	[56.74,62.26]	59.50

Table: Résultats obtenus avec les méthodes SVM et KNN testées sur les données obtenues par 2-means

Réseaux de neurones récurrents (RNN)

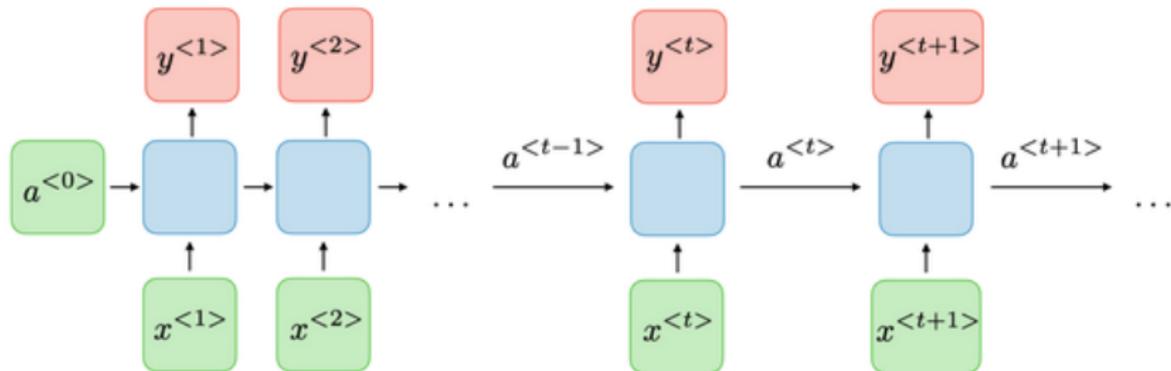


Figure: Réseau de neurones récurrent [8]

Réseau longue mémoire à court terme (LSTM)

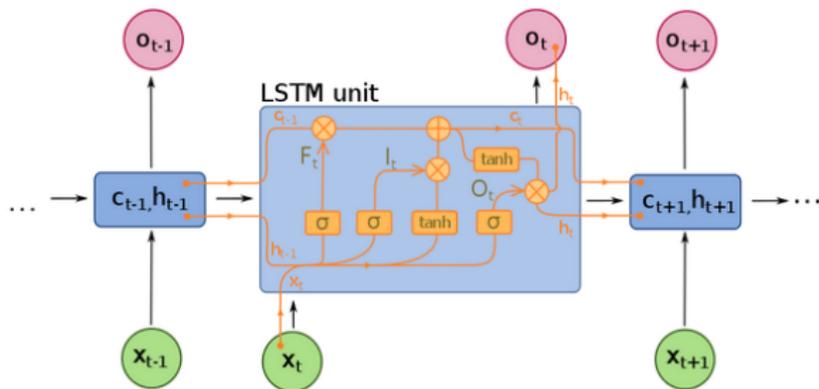


Figure: Cellule LSTM [9]

- ### LSTM
- Méthodes de Deep Learning
 - La solution à la mémoire à court terme
 - Possède des portes qui régularisent le flux d'information

Méthode LSTM appliquée sur la première base

Les paramètres utilisés sont :

- un modèle séquentiel avec une couche LSTM en entrée contenant 62 neurones et 2 neurone en sortie
- fonction d'activation : softmax
- dropout = 0.5
- batch-size = 10
- nombre d'épochs = 50
- optimiseur : Adam avec lr=0.01
- fonction de perte : la précision

Méthode LSTM appliquée sur la deuxième base

Les paramètres utilisés sont :

- un modèle séquentiel avec une couche LSTM en entrée contenant 62 neurones et 2 neurone en sortie
- fonction d'activation : softmax
- dropout = 0.2
- batch-size = 80
- nombre d'épochs = 100
- optimiseur : Adam avec lr=0.001
- fonction de perte : la précision

Méthode LSTM appliquée sur la deuxième base

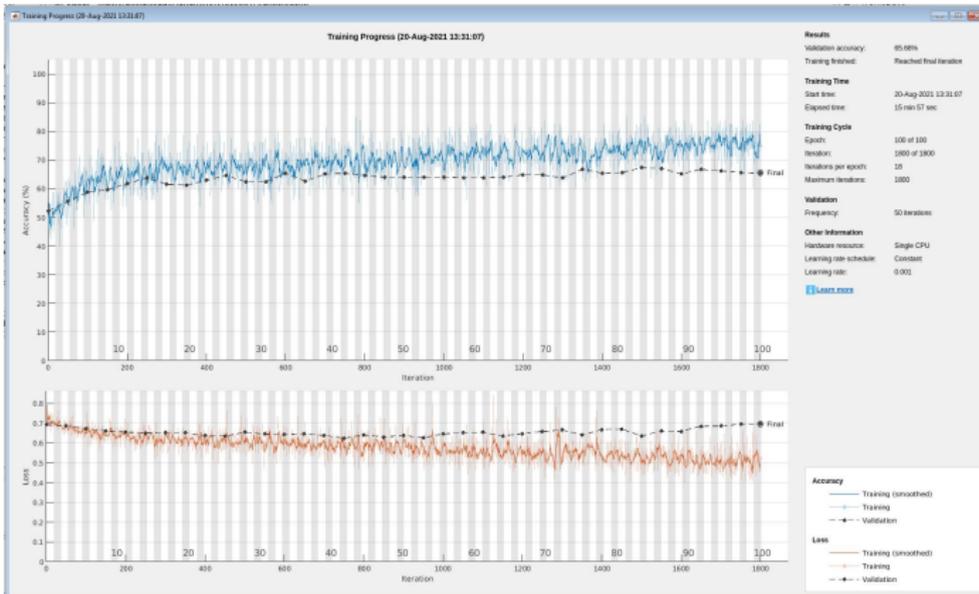


Figure: Représentation de la fonction de perte et de la précision des ensembles de validation et de test de la deuxième base

Méthode CNN appliquée sur la deuxième base

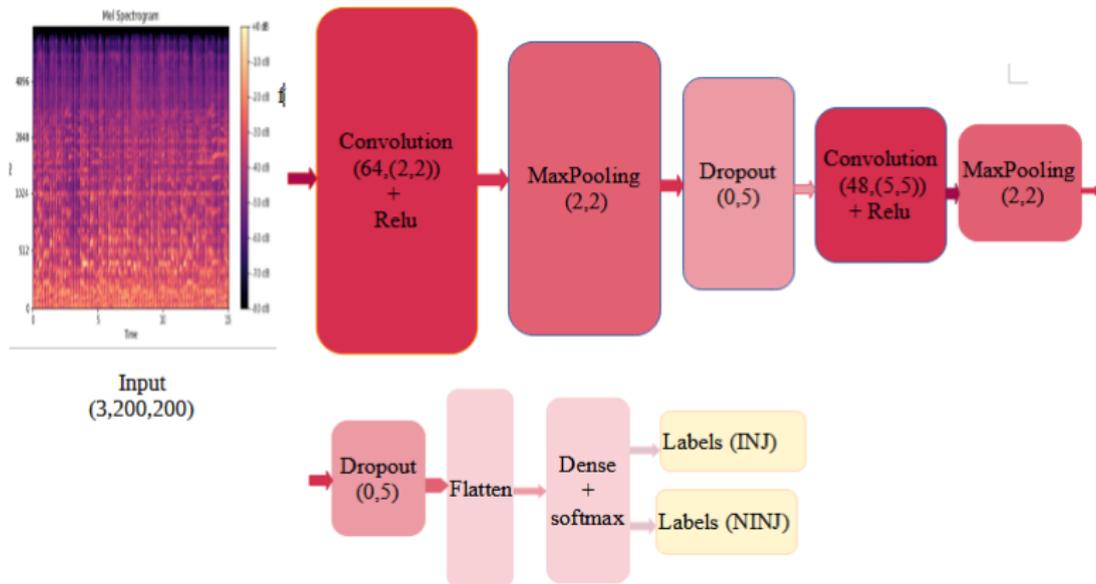


Figure: Architecture du CNN proposée

Méthode CNN appliquée sur la deuxième base



Figure: Représentation de la fonction de perte et de la précision des ensembles de validation et de test de la deuxième base

Conclusion

Conclusion

- Confirmation des résultats obtenus avec Hocine-Gharbi et al sur la première base
- La deuxième base contient toujours des audio non exploitables
- La deuxième base donne des meilleurs taux de classification en utilisant tous les caractéristiques des audio
- La méthode LSTM a amélioré les TC obtenus par les méthodes SVM et KNN avec la première base
- La méthode CNN améliore les TC de classification de la deuxième base

Perspectives

- Proposer d'autres architectures pour le CNN
- Tester sur une nouvelle base mieux structurée
- Utiliser hybridation CNN+LSTM
- Extraire d'autres caractéristique : MFCC...

Bibliographie



A Hacine-Gharbi et Ravier P. "Automatic Classification of French Spontaneous Oral Speech into Injunction and No-Injunction Classes". 2020



Comprendre le spectrogramme Mel <https://ichi.pro/fr/comprendre-le-spectrogramme-mel-277775661583955>



SVM <https://dataanalyticspost.com/Lexique/svm/>



c.fairy. Les K plus proches voisins. <https://htk.eng.cam.ac.uk/>



Validation croisée K-Fold pour le Deep Learning à l'aide de Keras. <https://ichi.pro/fr/validation-croisee-k-fold-pour-le-deep-learning-a-l-aide-de-keras-69014279685432>.



Wannous.H, Multi view classification of color regions application of the 3D assessment of chronic wounds. 2008



Marie-Jeanne Vieille.V. k-means, comment ça marche?, 2017



Afshine.A et Shervine.A. Pense-bête de réseaux de neurones récurrents



Comment le LSTM améliore le RNN <https://ichi.pro/fr/comment-le-lstm-ameliore-le-rnn-34021890806049>



dshahid380,Convolutional
NeuralNetwork, <https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529,2019>

Merci pour votre attention

