A FFNN Models in Details

This section discusses the details of the FFNN models.

A.1 Basic Model

After a massive exploration for finding the best hyperparameters to structure the model (like the number of layers, number of neurons in each layer, and the activation function), the final structure is shown in Table 11. This model results in 92.87%, 90.72% and 90.67% accuracies for training, validation, and testing datasets, respectively. Figure 7 shows the loss and accuracy values on the training and validation datasets while training. The model is still able to slightly learn as well as generalize even after 300 epochs with no signs of overfitting.

Table 11: FFNN basic model structure

Layer Name	Neurons	Activation Func
Hidden 1	200	ReLU
Hidden 2	500	ReLU
Hidden 3	500	ReLU
Hidden 4	450	ReLU
Hidden 5	400	ReLU
Hidden 6	400	ReLU
Hidden 7	350	ReLU
Hidden 8	300	ReLU
Hidden 9	300	ReLU
Hidden 10	250	ReLU
Hidden 11	200	ReLU
Hidden 12	200	ReLU
Hidden 13	150	ReLU
Hidden 14	100	ReLU
Hidden 15	100	ReLU
Hidden 16	50	ReLU
Hidden 17	25	ReLU
Output	15	Softmax
Trainable Parameters: 1,501,115		

A.2 100-Hot Model

Starting from the the basic model structure (Table 11), and by tuning the hyperparameters and using extra techniques, such as applying Dropout regularization, the model is able to learn better using the 100-hot representations. The model is structured as shown in Table 12. This model results in 94.25%, 93.49% and 93.45% accuracies for training, validation, and testing datasets, respectively. This is an improvement of 2.78% on the test set accuracy compared to the basic model. Figure 8 shows the loss and accuracy values on the training and validation datasets while training.

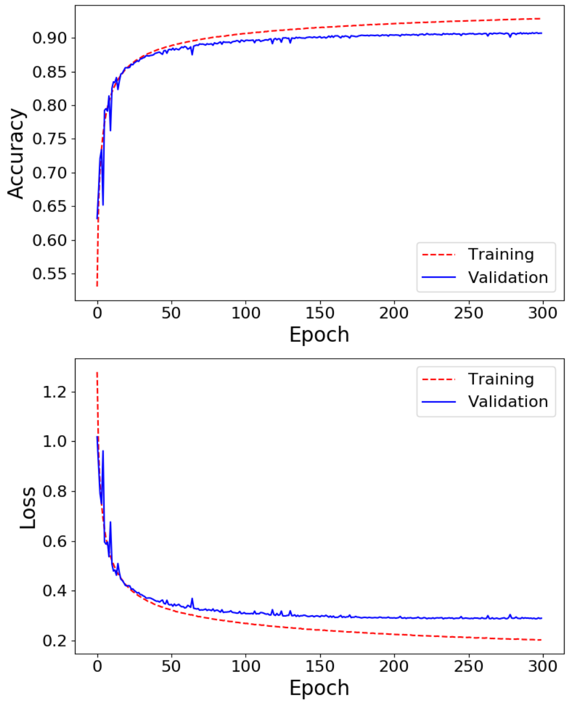

Figure 7: FFNN basic model training and validation accuracy and loss.

A.3 Embeddings Model

Using a very similar structure as the 100-hot model structure (Table 12), this model is structured as shown in Table 13. It achieves the best results compared to the basic and 100-hot models with 94.88%, 94.53% and 94.49% accuracies for training, validation, and testing datasets, respectively. This model improves the accuracy by 1.04% on the test set compared to the 100-hot model while reducing the number of trainable parameters (model size) by 51.46% and 62.66% compared to the basic and 100-hot models, respectively. Figure 9 shows the loss and accuracy values on the training and validation datasets while training.

Figure 10 shows the best diacritization examples diacritized using each FFNN model, while Figure 11 shows the worst diacritization examples. It is worth mentioning that the worst examples (listed in Figure 11) are from old Arabic poetry, which is very hard to diacritize flawlessly even for native speakers.

Table 12: FFNN 100-Hot model structure

Layer Name	Neurons	Activation Func
One Hot	N/A	N/A
Flatten	N/A	N/A
Dropout 1 (2.5%)	N/A	N/A
Hidden 1	250	ReLU
Dropout 2 (2.5%)	N/A	N/A
Hidden 2	200	ReLU
Dropout 3 (2.5%)	N/A	N/A
Hidden 3	150	ReLU
Dropout 4 (2.5%)	N/A	N/A
Hidden 4	100	ReLU
Dropout 5 (2.5%)	N/A	N/A
Hidden 5	50	ReLU
Dropout 6 (2.5%)	N/A	N/A
Output	15	Softmax
Trainable Parameters: 1,951,515		

Table 13: FFNN Embeddings model structure

Layer Name	Neurons	Activation Func
Embedding (25)	N/A	N/A
Flatten	N/A	N/A
Dropout (10%)	N/A	N/A
Hidden 1	250	ReLU
Hidden 2	200	ReLU
Hidden 3	150	ReLU
Hidden 4	100	ReLU
Hidden 5	50	ReLU
Output	15	Softmax
Trainable Parameters: 728,590		

B RNN Models in Details

This section provides details for the trained RNN models. First of all, Figure 12 shows the validation DER of each model while training, reported every 5 epochs. This clarifies the importance of the dataset size, where any model significantly improves their DER when trained with the extra train dataset compared to any other model trained without it.

Moreover, to explore the embeddings learnt by our best model, the weights vectors from the embeddings layer were extracted and reduced to 2 dimensions instead of 25 using t-SNE dimensionality reduction algorithm (Maaten and Hinton, 2008), then plotted in 2D space as shown in Figure 13. The embeddings are able to capture meaningful information where digits appear together at the bottom-left, the majority of the punctuations appear at the middle and the top-left side, and finally, the Arabic letters appear at the right side.

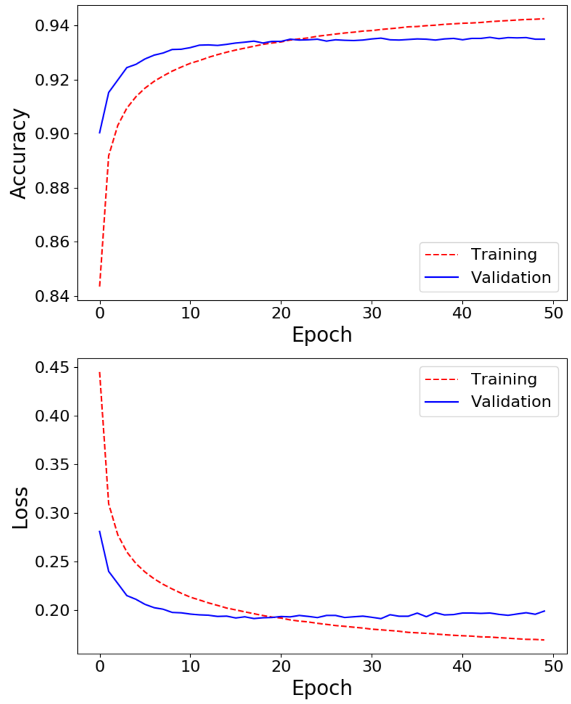

Figure 8: FFNN 100-Hot model training and validation accuracy and loss.

Figures 14 and 15 show both best and worst examples from diacritizing using each RNN model. An important note is that the old Arabic poetry lines are no longer the majority in the worst examples, in contrast to the FFNN models.

Finally, Figures 16 and 17 shows the confusion matrices related to our best model when trained without and with the extra train dataset, respectively. By comparing them, it is easy to see that the Shadda class is the worst one in both cases. However, the case with the extra train dataset shows dramatic improvement in this class, as well as other classes like Shadda + another diacritic and the Dammatan. A justification for this improvement is that there is a larger number of examples in the extra train dataset related to these classes as shown in Table 14. Another insight can be concluded from the confusion matrices is that the model usually misclassifies the Shadda class as Shadda + another diacritic class due to different diacritization conventions, which in many cases would be a grammatically correct guess.

Table 14: Number of examples for each class

	Train	Valid	Test	Extra Train	Total	%
No Diacritic	4,366K	213K	222K	46,647K	51,449K	38.87
Fatha	2,932K	144K	150K	31,287K	34,514K	26.07
Fathatah	58K	3K	3K	626K	691K	00.52
Damma	812K	39K	41K	8,648K	9,539K	07.20
Dammatan	58K	3K	3K	622K	686K	00.51
Kasra	1,265K	62K	64K	13,533K	14,924K	11.27
Kasratan	88K	4K	4K	941K	1,037K	00.78
Sukun	1,230K	60K	63K	13,135K	14,487K	10.94
Shaddah	6K	254	471	66K	73K	00.05
Shaddah + Fatha	300K	15K	15K	3,202K	3,532K	02.66
Shaddah + Fathatah	3K	189	132	36K	40K	00.03
Shaddah + Damma	43K	2K	2K	463K	511K	00.38
Shaddah + Dammatan	5K	238	222	51K	56K	00.04
Shaddah + Kasra	64K	3K	3K	679K	749K	00.56
Shaddah + Kasratan	6K	298	273	63K	69K	00.05

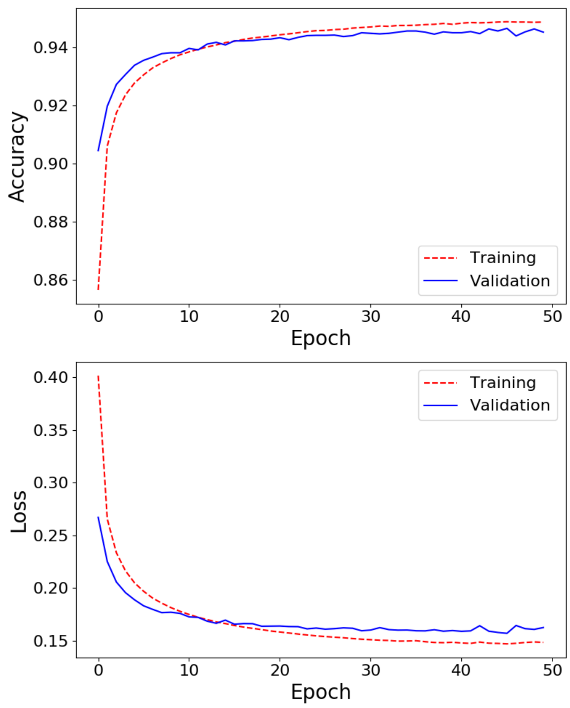

Figure 9: FFNN Embeddings model training and validation accuracy and loss.

Model		Best Line	File
FFNN Basic	Correct Diacritization	464 - حَدَّثَنِي يَحْيَى عَنْ مَالِك عَنْ هِشَامِ بْنِ عُرْوَةَ عَنْ أَبِيهِ Buckwalter Transliteration: 464 - Had~avaniy yaHoyaY Eano maAlik Eano hi$aAmi boni Eurowapa Eano >abiyhi	موطأ مالك
	Model's Output	464 - حَدَّثَنِي يَحْيَى عَنْ مَالِك عَنْ هِشَامِ بْنِ عُرْوَةَ عَنْ أَبِيهِ Buckwalter Transliteration: 464 - Had~avaniy yaHoyaY Eano maAlik Eano hi$aAmi boni Eurowapa Eano >abiyhi	
FFNN 100-Hot	Correct Diacritization	فَإِذَا ادَّعَى شَخْصٌ عَلَى غَيْرِهِ بِأَنَّهُ رَقِيقٌ فَعَلَيْهِ الْبَيَانُ . Buckwalter Transliteration: fa<i*aA Ad~aEaY $axoSN EalaY gayorihi bi>an~ahu raqiyqN faEalayohi AlobayaAnu .	الروض الأنف
	Model's Output	فَإِذَا ادَّعَى شَخْصٌ عَلَى غَيْرِهِ بِأَنَّهُ رَقِيقٌ فَعَلَيْهِ الْبَيَانُ . Buckwalter Transliteration: fa<i*aA Ad~aEaY $axoSN EalaY gayorihi bi>an~ahu raqiyqN faEalayohi AlobayaAnu .	
FFNN Embeddings	Correct Diacritization	وَالْهَاوَنُ مِثَالٌ ، فَمِثْلُهُ كُلُّ مَا يَتَعَذَّرُ كَسْرُهُ عَلَى رَأْسِهَا . Buckwalter Transliteration: waAlohaAwanu mivaAlN ، famivoluhu kul~u maA yataEa*~aru kasoruhu EalaY ra>osihaA .	مبعث النبي صلى الله عليه وسلم
	Model's Output	وَالْهَاوَنُ مِثَالٌ ، فَمِثْلُهُ كُلُّ مَا يَتَعَذَّرُ كَسْرُهُ عَلَى رَأْسِهَا . Buckwalter Transliteration: waAlohaAwanu mivaAlN ، famivoluhu kul~u maA yataEa*~aru kasoruhu EalaY ra>osihaA .	

Figure 10: FFNN models good diacritization examples.

Model		Worst Line	File
FFNN Basic	Correct Diacritization	أَنْصَابِ مَكَّةَ عَامِدِينَ لِيَثْرِبِ ... فِي ذِي غَيَاطِلَ جَحْفَلٍ جَبْجَابِ Buckwalter Transliteration: >anoSaAbi mak~pa EaAmidiyna liyavoribi ... fiy *iy gayaATila jaHofalK jabojaAbi	سيرة ابن هشام
	Model's Output	أنْصَابَ مَكَّةَ عَامِدِّيْنِ لِيُثِرِبُ ... فِي ذِي غِيَاطُلُ جَحَفَلُ جَبْجَابٍ Buckwalter Transliteration: >noSaAba mak~apa EaAmid~iyoni liyuviribu ... fiy *iy giyaATulu jaHafalu jabojaAbK	
FFNN 100-Hot	Correct Diacritization	لَوْلَا جَرِيرٌ هَلَكَتْ بَجِيلَهْ ... نِعْمَ الْفَتَى ، وَبِئْسَ الْقَبِيلَهْ Buckwalter Transliteration: lawolaA jariyrN halakato bajiylaho ... niEoma AlofataY ، wabi}osa Aloqabiylaho	الروض الأنف
	Model's Output	لَوَلَا جَرِيرَ هَلَكَتْ بِجَيْلِهِ ... نَعَمْ الْفَتَى ، وَبِئِسَ الْقَبِيلِهْ Buckwalter Transliteration: lawalaA jariyra halakato bijayolihi ... naEamo AlofataY ، wabi}isa Aloqabiyliho	
FFNN Embeddings	Correct Diacritization	فَكُلّ صَدِيقٍ وَابْنِ أُخْتٍ نَعُدّهُ لَعَمْرِي وَجَدْنَا غِبّهُ غَيْرَ طَائِلِ Buckwalter Transliteration: fakul~ SadiyqK waAboni >uxotK naEud~hu laEamoriy wajadonaA gib~hu gayora TaA}ili	مبعث النبي صلى الله عليه وسلم
	Model's Output	فَكُلُّ صَدِيقٍ وَابْنُ أُخْتِ نَعْدُهُ لِعُمْرِي وَجَدْنَا غَبْهِ غَيْرُ طَائِلٍ Buckwalter Transliteration: fakul~u SadiyqK waAbonu >uxoti naEoduhu liEumoriy wajadonaA gabohi gayoru TaA}ilK	

Figure 11: FFNN models bad diacritization examples.

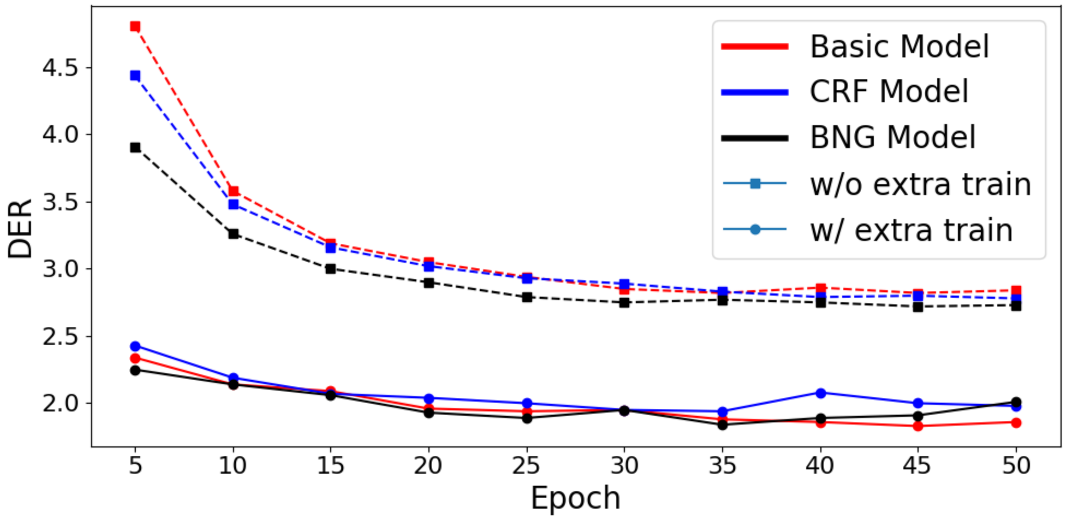

Figure 12: Recurrent models validation DER while training.

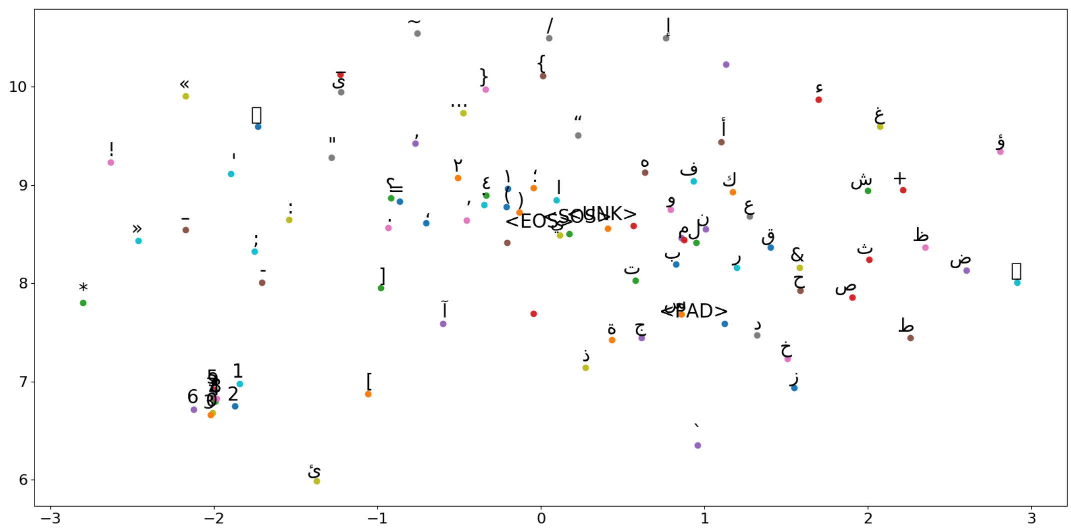

Figure 13: Embeddings plotted in 2D space.

Model		Best Line	File
RNN Basic	Correct Diacritization	قَوْلُهُ : (وَبَحَثَ الرَّافِعِيُّ صِحَّتَهَا) وَإِنْ قَصَدَ تَمْلِيكَ الْمَسْجِدِ وَهُوَ الْمُعْتَمَدُ Buckwalter Transliteration: qawoluhu : (wabaHava Alr~aAfiEiy~u SiH~atahaA) wa<ino qaSada tamoliyka Alomasojidi wahuwa AlomuEotamadu	حاشية البجيرمي على الخطيب
	Model's Output	قَوْلُهُ : (وَبَحَثَ الرَّافِعِيُّ صِحَّتَهَا) وَإِنْ قَصَدَ تَمْلِيكَ الْمَسْجِدِ وَهُوَ الْمُعْتَمَدُ Buckwalter Transliteration: lilomurotahini qabola Huluwli >ajali Ald~ayoni bi>ano qay~adahaA bizamanK >awo EamalK yanoqaDiy qabolahu	
RNN CRF	Correct Diacritization	لِلْمُرْتَهِنِ قَبْلَ حُلُولِ أَجَلِ الدَّيْنِ بِأَنْ قَيَّدَهَا بِزَمَنٍ أَوْ عَمَلٍ يَنْقَضِي قَبْلَهُ Buckwalter Transliteration: lilomurotahini qabola Huluwli >ajali Ald~ayoni bi>ano qay~adahaA bizamanK >awo EamalK yanoqaDiy qabolahu	منح الجليل شرح مختصر خليل
	Model's Output	لِلْمُرْتَهِنِ قَبْلَ حُلُولِ أَجَلِ الدَّيْنِ بِأَنْ قَيَّدَهَا بِزَمَنٍ أَوْ عَمَلٍ يَنْقَضِي قَبْلَهُ Buckwalter Transliteration: qawoluhu : (wabaHava Alr~aAfiEiy~u SiH~atahaA) wa<ino qaSada tamoliyka Alomasojidi wahuwa AlomuEotamadu	
RNN BNG	Correct Diacritization	وَلِذَا قَالَ وَلَوْ لَمْ يَقُلْ أَيْ الْمُوَثِّقُ إنْ فَعَلَ شَيْئًا مِنْهَا بِأَنْ قَالَ Buckwalter Transliteration: wali*aA qaAla walawo lamo yaqulo >ayo Alomuwav~iqu <no faEala $ayo}FA minohaA bi>ano qaAla	شرح مختصر خليل للخرشي
	Model's Output	وَلِذَا قَالَ وَلَوْ لَمْ يَقُلْ أَيْ الْمُوَثِّقُ إنْ فَعَلَ شَيْئًا مِنْهَا بِأَنْ قَالَ Buckwalter Transliteration: wali*aA qaAla walawo lamo yaqulo >ayo Alomuwav~iqu <no faEala $ayo}FA minohaA bi>ano qaAla	

Figure 14: RNN models good diacritization examples.

Model		Worst Line	File
RNN Basic	Correct Diacritization	وَبَكَّيهِ لِلْأَيْتَامِ وَالرِّيحُ زَفْزَفَةٌ ... وَتَشْبِيبُ قِدْرٍ طَالَمَا أَزْبَدَتْ تَغْلِي Buckwalter Transliteration: wabak~yhi lilo>ayotaAmi waAlr~yHu zafozapN ... wata$obiybu qidorK TaAlamaA >azobadato tagoliy	سيرة ابن هشام
	Model's Output	وَبَكَيْهُ لِلْأَيْتَامِ وَالرِّيحِ زَفْزَفَةُ ... وَتَشْبِيبُ قَدْرٍ طَالِمًا أَزْبَدْتْ تَغْلِي Buckwalter Transliteration: wabakayohu lilo>ayotaAmi waAlr~iyHi zafozapu ... wata$obiybu qadorK TaAlimFA >azobadoto tagoliy	
RNN CRF	Correct Diacritization	وَنَقَلَ فِي مَوْضِعٍ آخَرَ تَعْلِيلَهُ بِأَنَّهُ لَا يُقْلَعُ النَّجَسَ لِلزُّوجَتِهِ . Buckwalter Transliteration: wanaqala fiy mawoDEK |xara taEoliylahu bi>an~ahu laA yuqolEu Aln~ajasa liluzuwjatihi .	شرح البهجة الوردية
	Model's Output	وَنُقِلَ فِي مَوْضِعٍ آخَرَ تَعْلِيلُهُ بِأَنَّهُ لَا يَقْلَعُ النَّجِسَ لِلزَّوْجَتَهِ . Buckwalter Transliteration: wanuqila fiy mawoDEK |xara taEoliyluhu bi>an~ahu laA yaqolaEu Aln~ajisa lilz~awojatahi .	
RNN BNG	Correct Diacritization	وَنَقَلَ فِي مَوْضِعٍ آخَرَ تَعْلِيلَهُ بِأَنَّهُ لَا يُقْلَعُ النَّجَسَ لِلزُّوجَتِهِ . Buckwalter Transliteration: wanaqala fiy mawoDEK |xara taEoliylahu bi>an~ahu laA yuqolEu Aln~ajasa liluzuwjatihi .	شرح البهجة الوردية
	Model's Output	وَنُقِلَ فِي مَوْضِعٍ آخَرَ تَعْلِيلُهُ بِأَنَّهُ لَا يَقْلَعُ النَّجِسَ لِلزَّوْجَتَهْ . Buckwalter Transliteration: wanuqila fiy mawoDEK |xara taEoliyluhu bi>an~ahu laA yaqolaEu Aln~ajisa lilz~awojataho .	

Figure 15: RNN models bad diacritization examples.

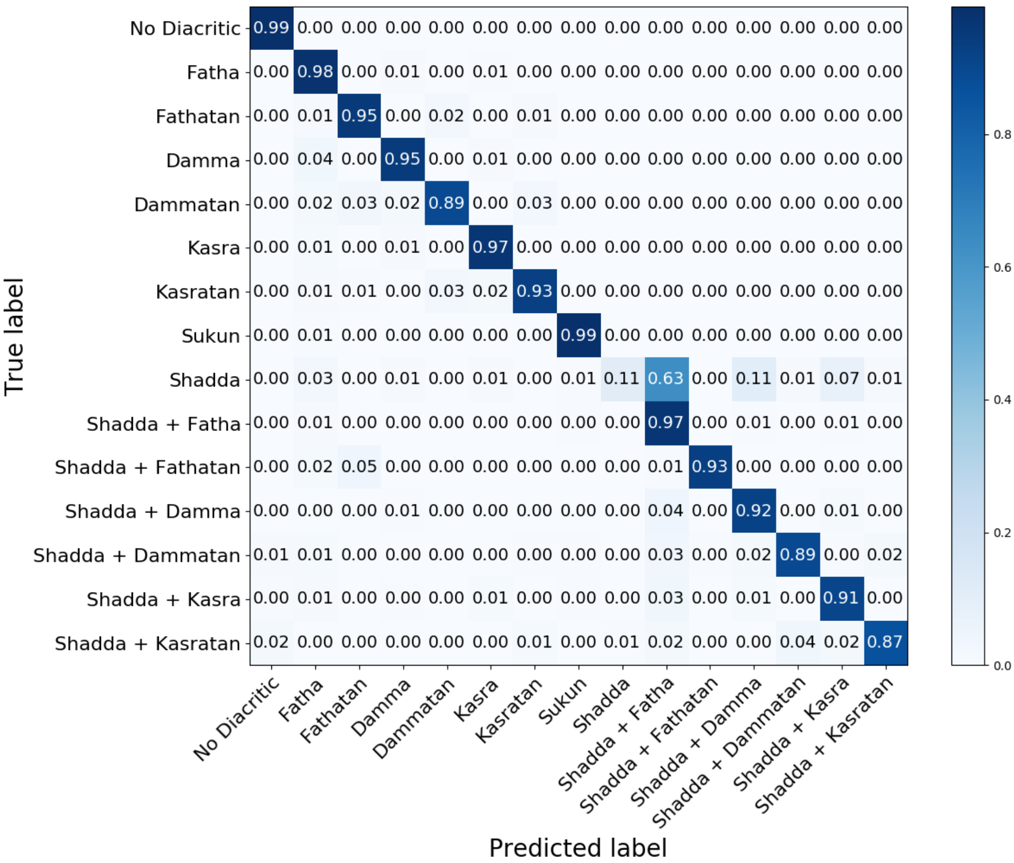

Figure 16: Without extra train confusion matrix for the best BNG model.

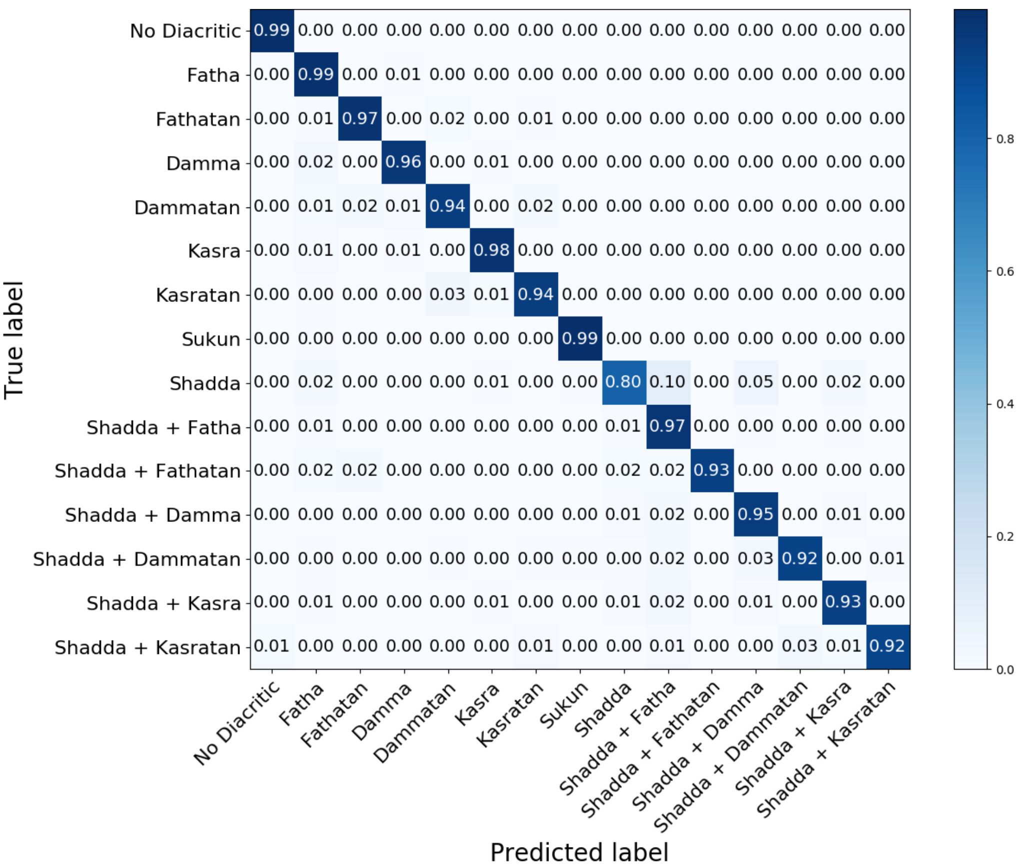

Figure 17: With extra train confusion matrix for the best BNG model.