Digital Data in Healthcare: Advanced Applications and Challenges

Presented By:
Dr. Hossam Faris
April 10th, 2021

Outline

- Telemedicine
- Altibbi
- Data Sources
- Pandemic effect
- Data Science @Altibbi
- AI/ML projects
- Challenges

Telemedicine

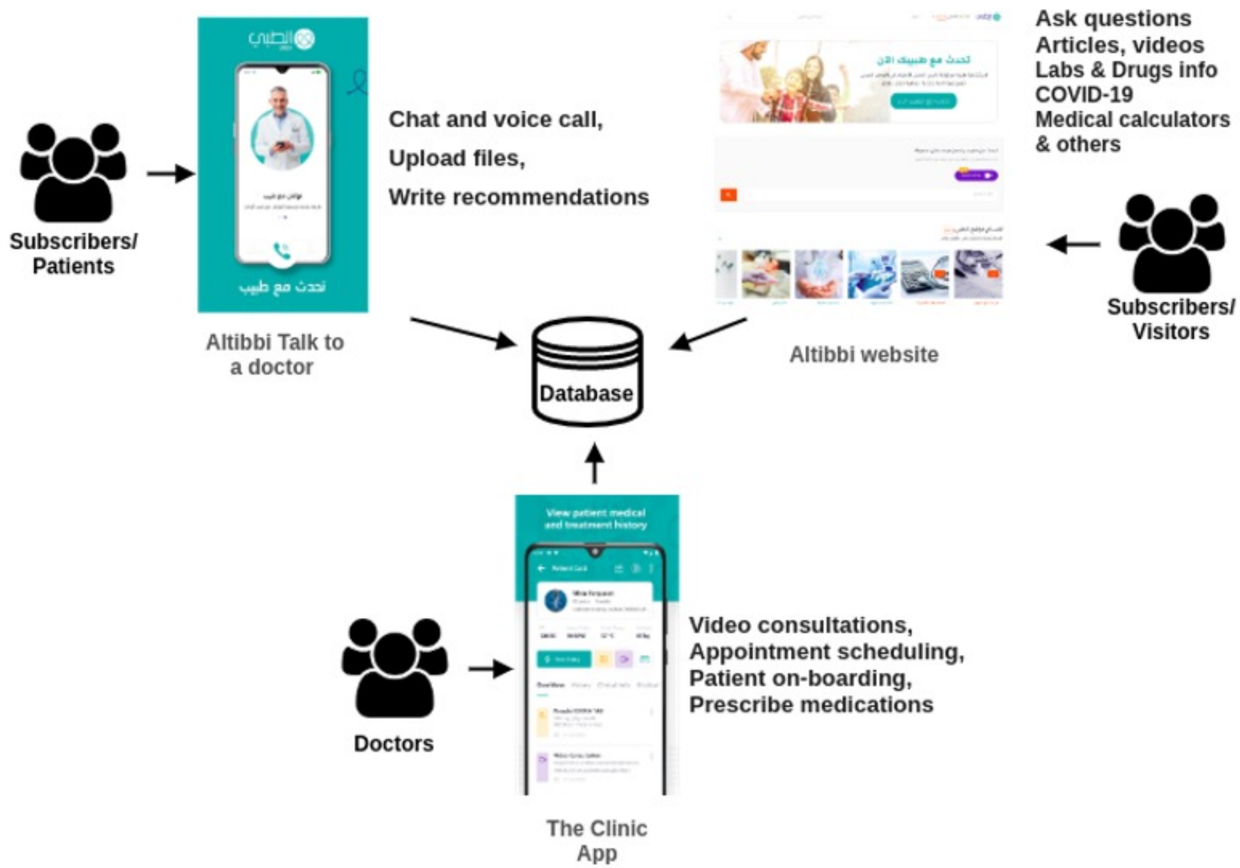
Telemedicine, a term coined in the 1970s.

Means “healing at a distance”, which signifies the use of ICT to improve patient outcomes by increasing access to care and medical information.

The context: Altibbi (www.altibbi.com)

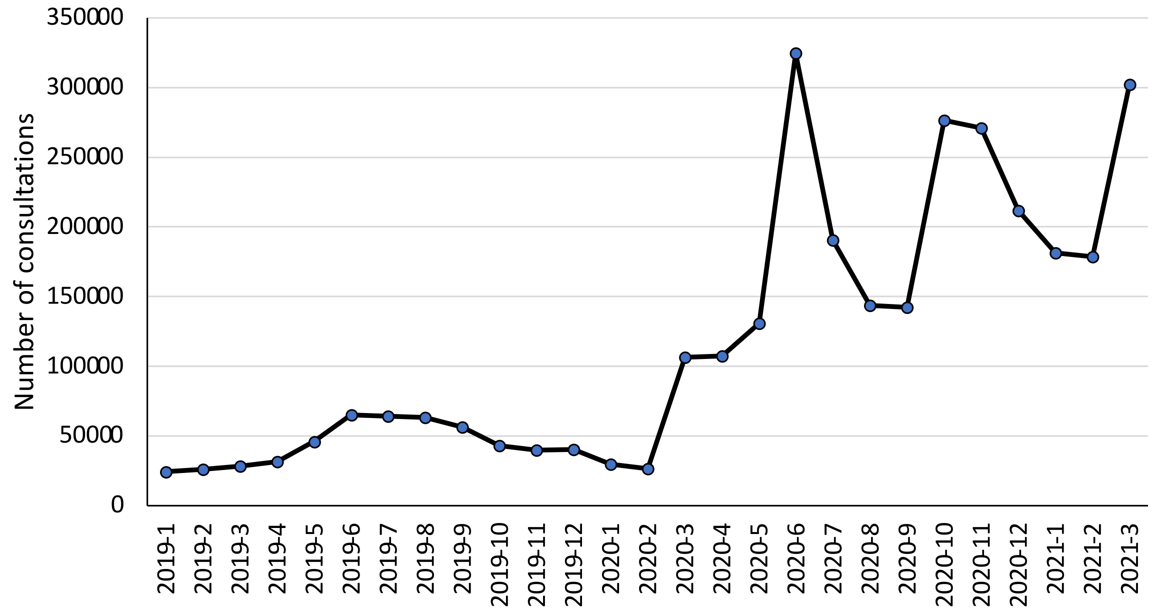
- It is one of the largest digital health platform in the Middle East and North Africa (**MENA**).
- Launched in **2011** in Amman, Jordan.
- The platform aims for presenting **telehealth services** & **simplified medical information** to users in the region in **Arabic**.
- Altibbi Telehealth increases people's access to quality **primary healthcare** while **lowering** the **risk of disease transmission.**

Sources of data at Altibbi



The effect of the pandemic

1. Altibbi is witnessing a huge demand for the service.
2. Altibbi covered **2 million consultations** from March 2020 till January 2021!



Altibbi

&

Data Science

The data science department has been officially structured (after ~**18** months) of preparation and building **skills** & **capacity**.

The **aim** is to improve the **quality** of different processes & services at different levels **using state-of-the-art** data science & **DL** technologies.

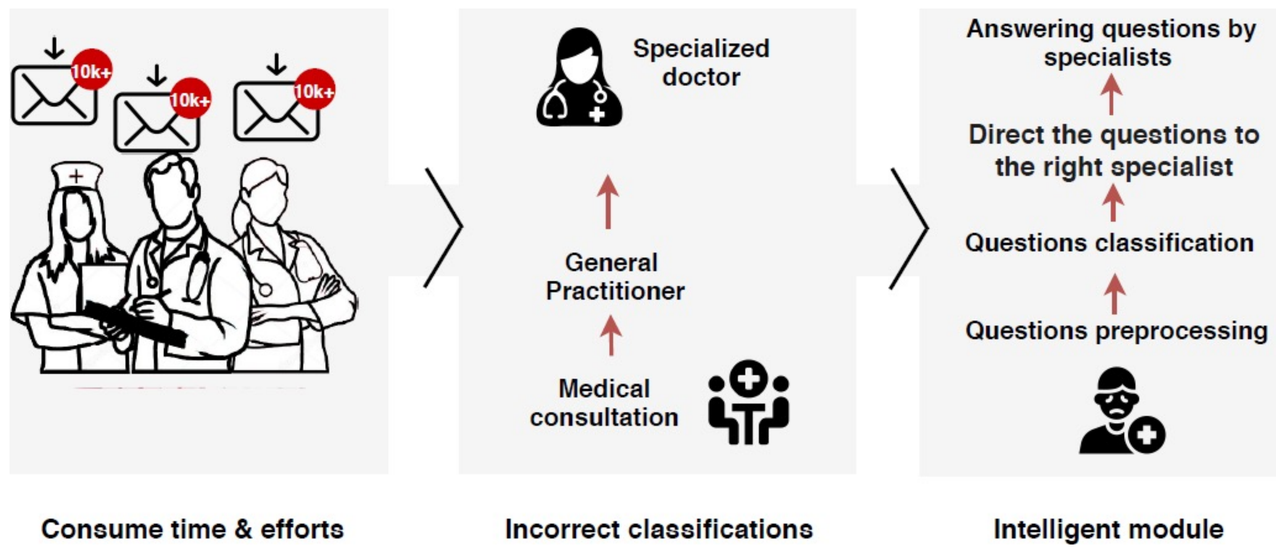
Altibbi Data Science Roadmap

- Supportive Diagnostic Module
- Supportive Treatment Module
- Automate Question Rejection
- Telemedicine Smart Routing
- Computer Vision Diagnosis
- iDoctor bot
- Free Q&A Classifier
- Specialty Detection
- Churn Prediction
- Question Answering
- Automate Quality Control of Consultations
- Content Cannibalization
- COVID-19 Calls Forecasting
- Dynamic/Personalized Content

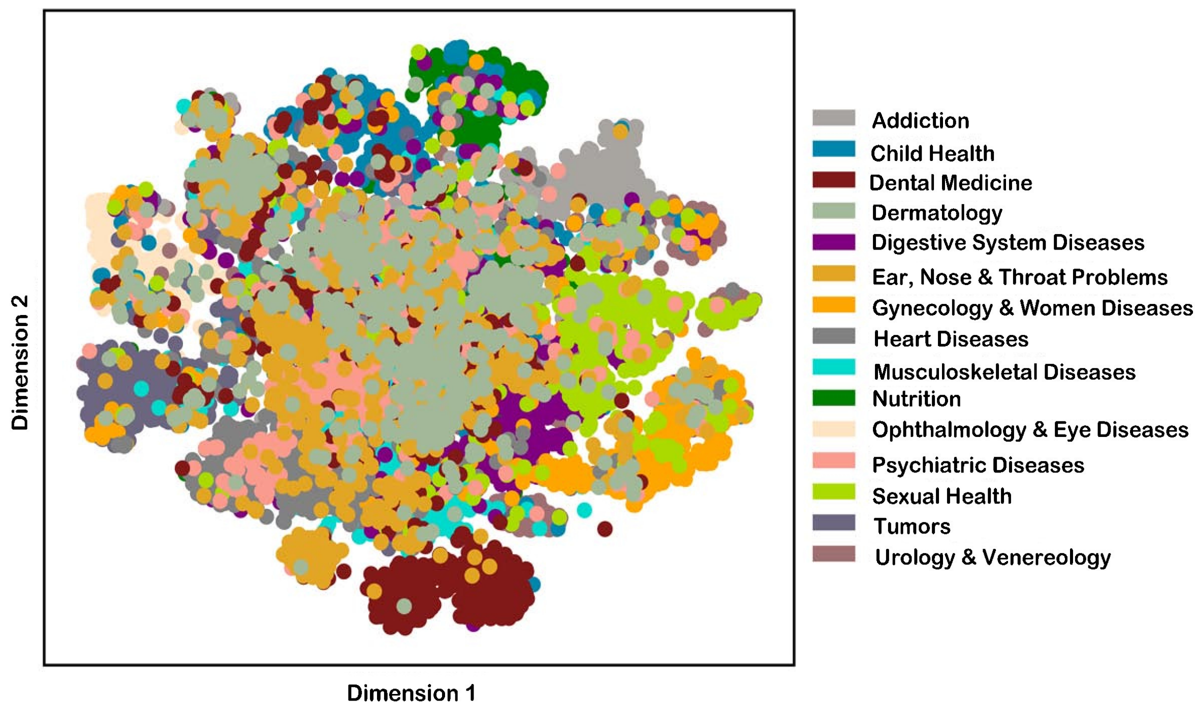
AI/ML-based Project

- Define Task
- Define data sources
- Exploratory Data Analysis (Visualize the dataset, summary statistics)
- Feature Engineering
- Predictive Modeling (e.g. DNN)
- Evaluation, Visualization, & Tuning

Specialty Detection (Problem and Proposed Module)



Specialty Detection
(Data Visualization)

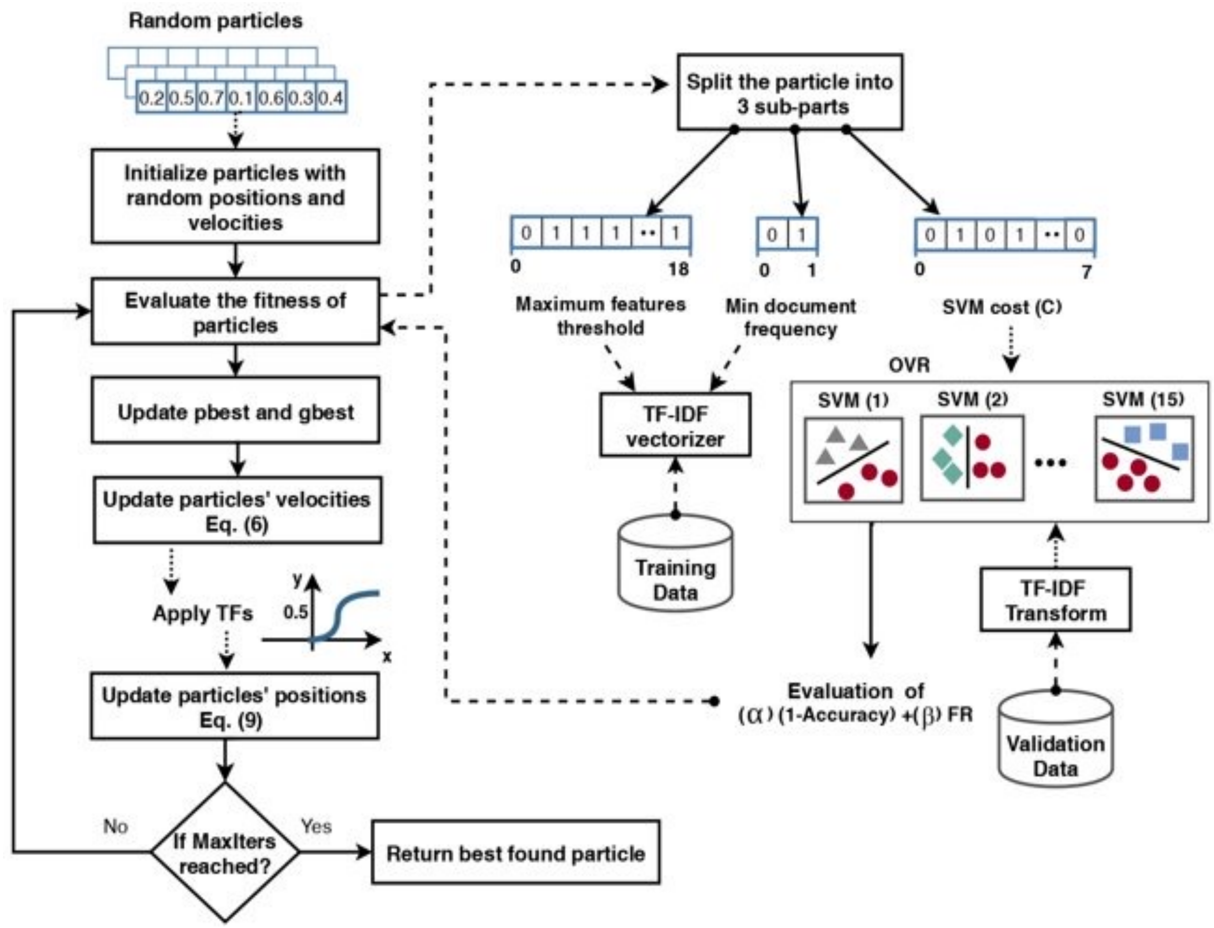


tSNE projection of 15,000 questions after preprocessing

Specialty Detection (A classical approach)

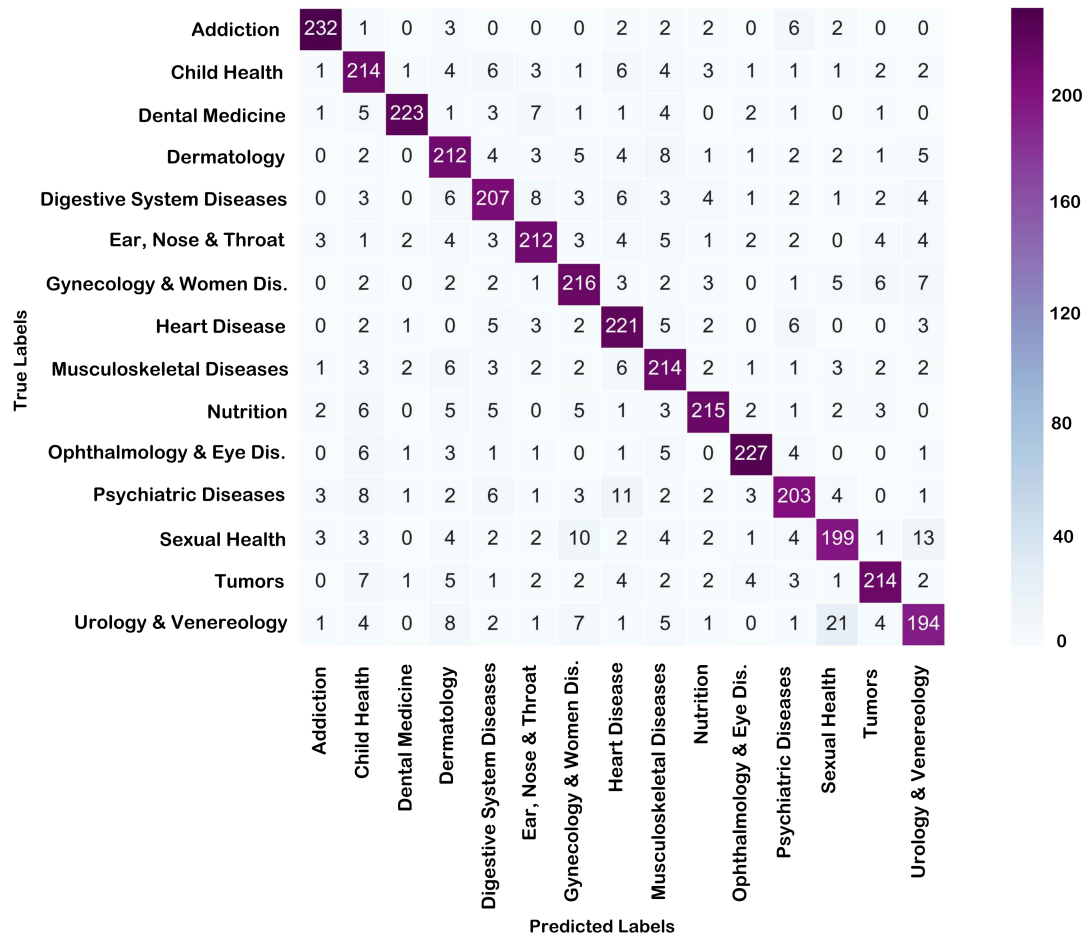
- Feature representation: **TF-IDF**
- Classifier: **SVM**
- Multi class strategy: **One-VS-Rest**

Specialty Detection (A classical approach)



Specialty Detection (Results)

Heatmap of the confusion matrix



Specialty Detection (Results)

Table 4: A comparison of the average performance of $BPSO_{S4} - SVM_{OVR}$ with other machine learning algorithms considering the number of all features. P-Values are based on ($\alpha = 0.05$), where the significant results are with underline typeface.

Algorithm	**Accuracy**	$F1 - score_m$	$Recall_m$	$Precision_m$
$BPSO_{S4} - SVM_{OVR}$	**0.852 ± 0.001**	**0.851 ± 0.001**	**0.851 ± 0.001**	**0.852 ± 0.001**
linearSVM(C=0.5)	0.844 ± 0.000 2.8719E-11	0.844 ± 0.000	0.844 ± 0.000	0.845 ± 0.000
Random Forest	0.736 ± 0.006 2.8719E-11	0.735 ± 0.007	0.736 ± 0.006	0.739 ± 0.007
Logistic Regression	0.835 ± 0.003 2.8719E-11	0.836 ± 0.003	0.839 ± 0.003	0.840 ± 0.003
MultinomialNB	0.812 ± 0.036 2.8719E-11	0.812 ± 0.036	0.812 ± 0.036	0.814 ± 0.035
ComplementNB	0.791 ± 0.023 2.8719E-11	0.789 ± 0.023	0.791 ± 0.023	0.791 ± 0.023
BernoulliNB	0.809 ± 0.033 2.8719E-11	0.809 ± 0.033	0.809 ± 0.033	0.811 ± 0.032
SGDClassifier	0.842 ± 0.001 2.8719E-11	0.842 ± 0.001	0.842 ± 0.001	0.842 ± 0.001
SVMs (RBF)	0.250 ± 0.280 2.8719E-11	0.230 ± 0.300	0.250 ± 0.280	0.750 ± 0.060
XGBoost	0.783 ± 0.010 2.8719E-11	0.783 ± 0.010	0.783 ± 0.010	0.785 ± 0.010
Adaboost	0.615 ± 0.005 2.8719E-11	0.626 ± 0.008	0.615 ± 0.005	0.686 ± 0.010
KNN	0.761 ± 0.042 2.8719E-11	0.761 ± 0.042	0.761 ± 0.042	0.765 ± 0.039

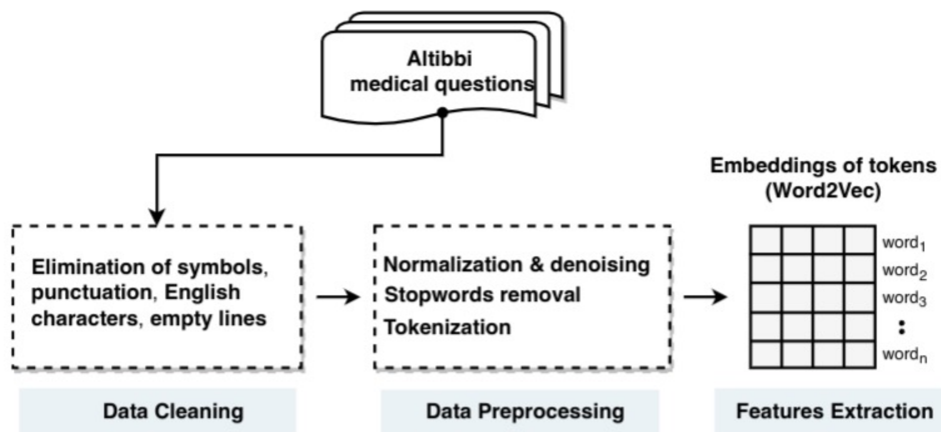
Specialty Detection (An attempt for improvement)

Targeted improvements:

- Feature representation level: Word Embeddings
- Classifier level: Deep learning

Specialty Detection: A deep learning approach (Pre-processing)

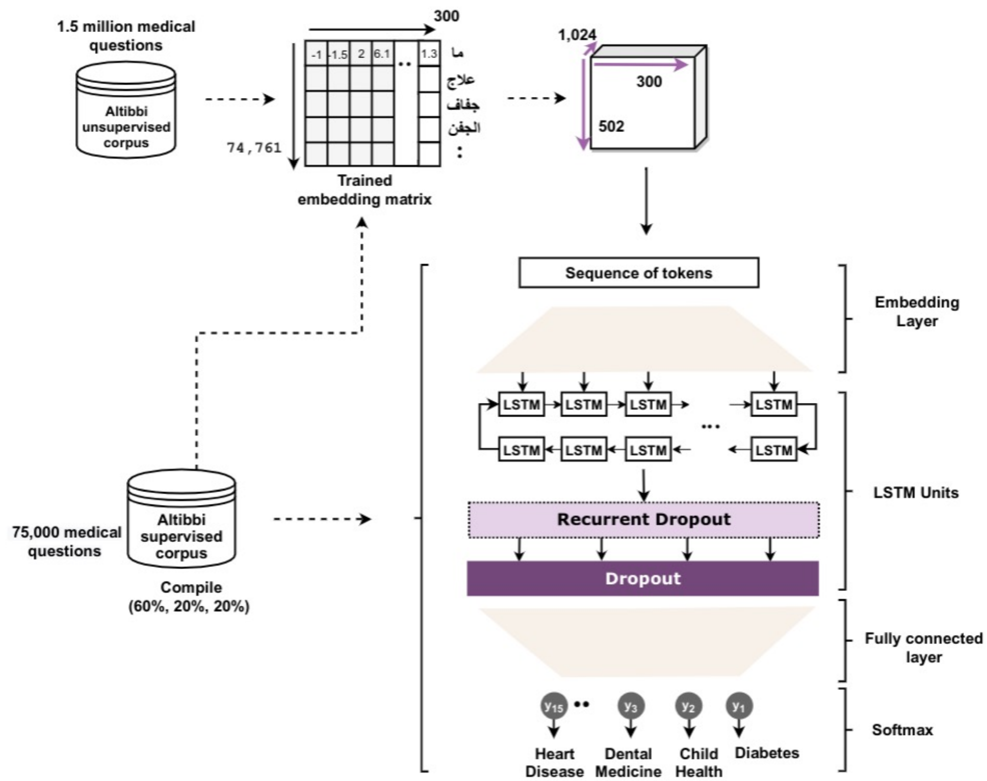
1.5 million unlabeled consultations



Specialty Detection: A deep learning approach (Feature representation)

- Word embedding representation based on **1.5 million** unlabeled consultations
- Types of experimented embeddings:
  - **Keras embedding**
  - **Aravec – Twitter** (Pre-Trained)
  - **Aravec – Wikipedia** (Pre-Trained)
  - **AltibbiVec**

Specialty Detection: A deep learning approach (Model)



Specialty Detection: A deep learning approach (results)

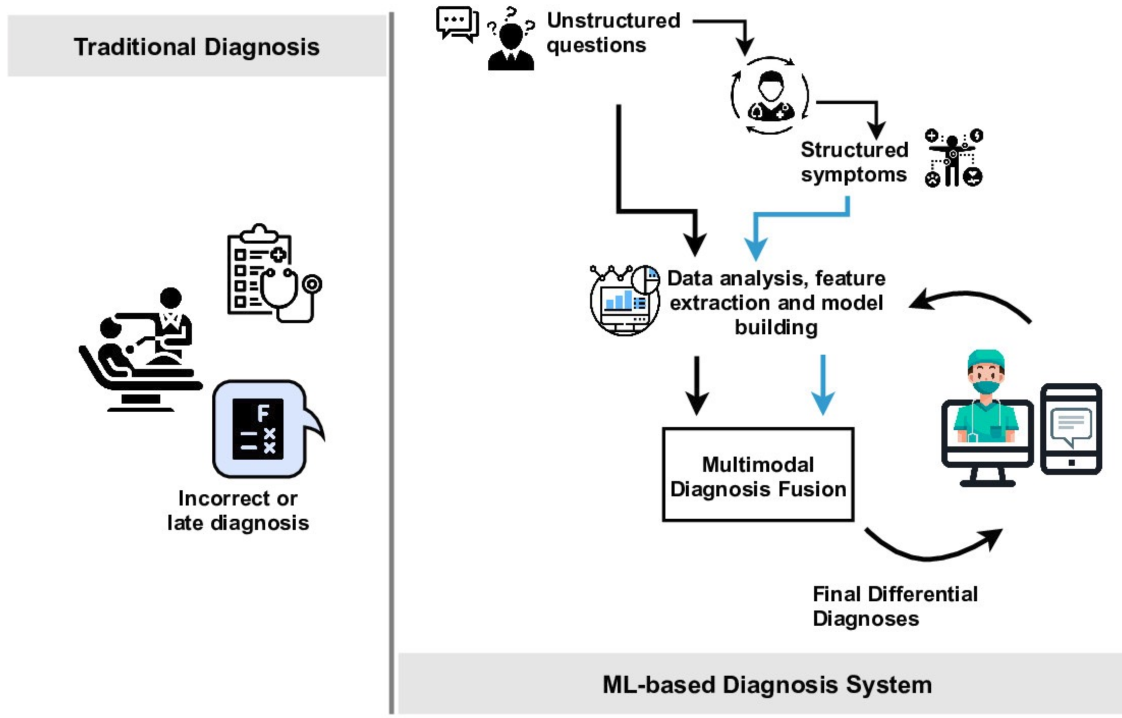
Table 4: Comparison between BiLSTM (30) and LSTM (40) based on precision, recall, and f1-scores for all classes.

	BiLSTM (30)			LSTM (40)		
Class	**Precision**	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**
Diabetes	**0.859**	0.854	**0.857**	0.850	**0.861**	0.855
Child Health	**0.949**	0.945	0.947	0.941	**0.965**	**0.953**
Ear, Nose & Throat	**0.844**	0.824	**0.834**	0.819	0.824	0.822
Dental Medicine	0.902	**0.922**	**0.912**	**0.909**	0.912	0.911
Nutrition	0.859	**0.847**	**0.853**	**0.860**	0.842	0.851
Ophthalmology & Eye Diseases	0.841	**0.869**	0.855	**0.868**	0.861	**0.864**
Dermatology	0.853	**0.888**	**0.870**	**0.859**	0.873	0.866
Heart Disease	**0.860**	0.873	0.867	0.855	**0.879**	0.867
Tumors	**0.873**	0.855	0.864	0.872	**0.862**	**0.867**
Psychiatric Diseases	**0.888**	0.860	**0.874**	0.865	**0.875**	0.870
Urology & Venereology	0.917	**0.953**	0.935	**0.922**	0.949	0.935
Digestive System Diseases	0.890	0.875	0.883	**0.902**	**0.880**	**0.891**
Musculoskeletal Diseases	0.833	**0.824**	**0.829**	**0.837**	0.813	0.825
Sexual Health	0.878	0.868	0.873	**0.880**	**0.870**	**0.875**
Gynecology & Women Diseases	**0.832**	**0.824**	**0.828**	0.827	0.803	0.815
Macro-average	**0.872**	**0.872**	**0.872**	0.871	0.871	0.871

Challenges in Specialty Detection

- The Arabic language is a morphologically rich and sophisticated language.
- The use of colloquial Arabic (mix of dialects).
- There are specialties that are extremely hard to correctly classify due to the small number of instances under these specialties. (imbalanced distribution)
- Some questions could be multi-labeled.
- Some specialties have questions that are very similar syntactically and semantically.

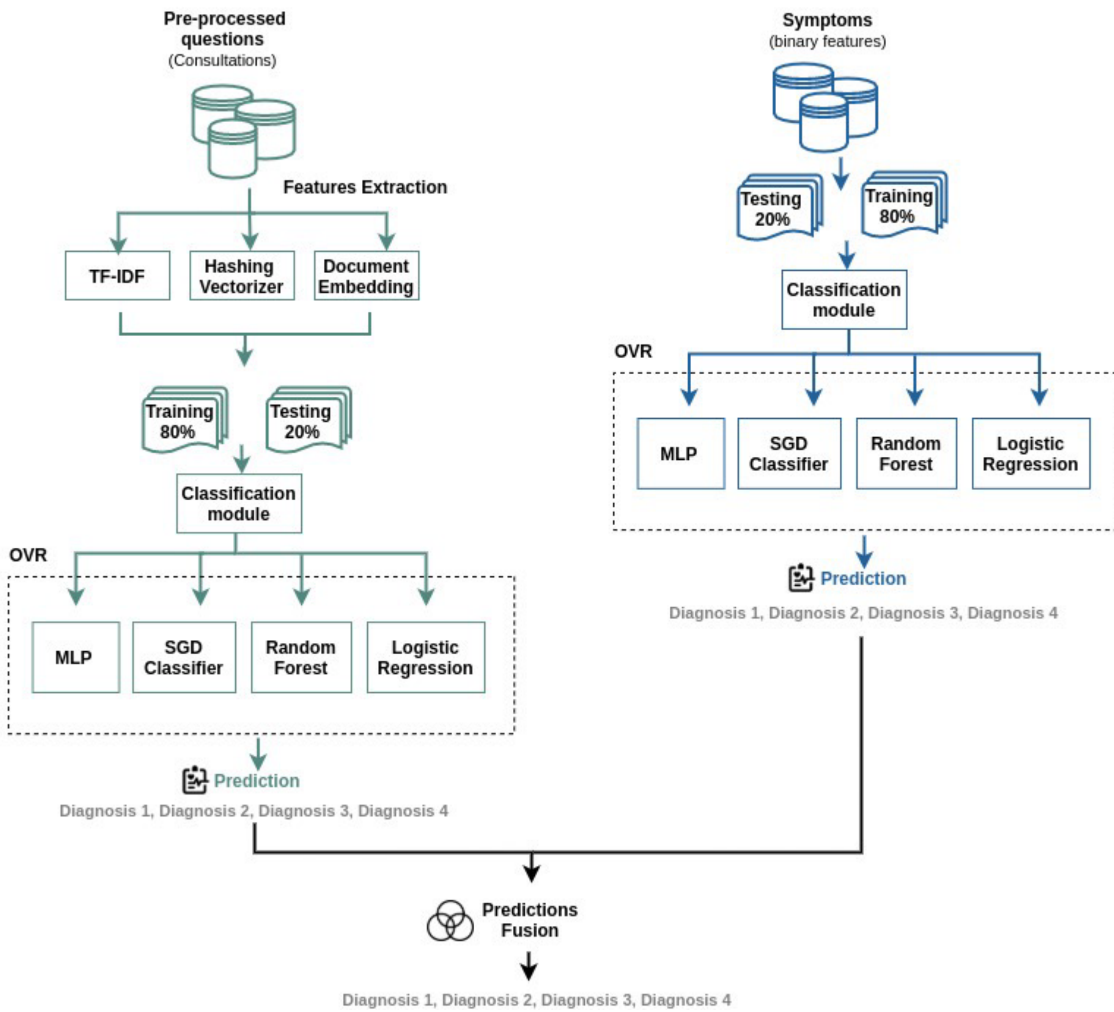
Supportive Diagnostic (Problem and Proposed Module)



Supportive Diagnostic Module (Data)

- The total collected data from Altibbi is **263,867** of questions (consultations) that are accompanied by symptoms and diagnoses.
- The total number of **symptoms** is **7,324**, while the **diagnoses** are **7,410**. Each consultation is accompanied by **multiple symptoms and multiple diagnoses** even that some of them infrequently occur.
- The diagnoses that are **repeated less than 20 times** were removed; the final number of **questions is 246,814,** and for the **diagnoses is 1206.**

SDM (Model)


Pre-processed questions
(Consultations)
Features Extraction
TF-IDF
Hashing Vectorizer
Document Embedding
Training 80%
Testing 20%
Classification module
OVR
MLP
SGD Classifier
Random Forest
Logistic Regression
Prediction
Diagnosis 1, Diagnosis 2, Diagnosis 3, Diagnosis 4
Symptoms
(binary features)
Testing 20%
Training 80%
Classification module
OVR
MLP
SGD Classifier
Random Forest
Logistic Regression
Prediction
Diagnosis 1, Diagnosis 2, Diagnosis 3, Diagnosis 4
Predictions Fusion
Diagnosis 1, Diagnosis 2, Diagnosis 3, Diagnosis 4

SDM (Results)

Table 6: The accuracy score of the final prediction based on four fusion criteria: the ranking of case I (Ranking-I), and of case II (Ranking-II), the summation, and multiplication.

	Accuracy			
	Ranking-I	**Ranking-II**	**Summation**	**Multiplication**
Precision_1	**0.813**	**0.828**	**0.846**	**0.849**
Precision_2	0.761	0.784	0.809	0.811
Precision_3	0.741	0.769	0.796	0.798

SDM (Challenges)

- Associations between diagnoses is not detected.
- Rare diagnoses are extremely hard to detect.

Quality Control of Medical Consultations over Voice (Problem Description)

The aim behind **this evaluation is to aid the medical operations team in recognizing** low-quality consultations**.**

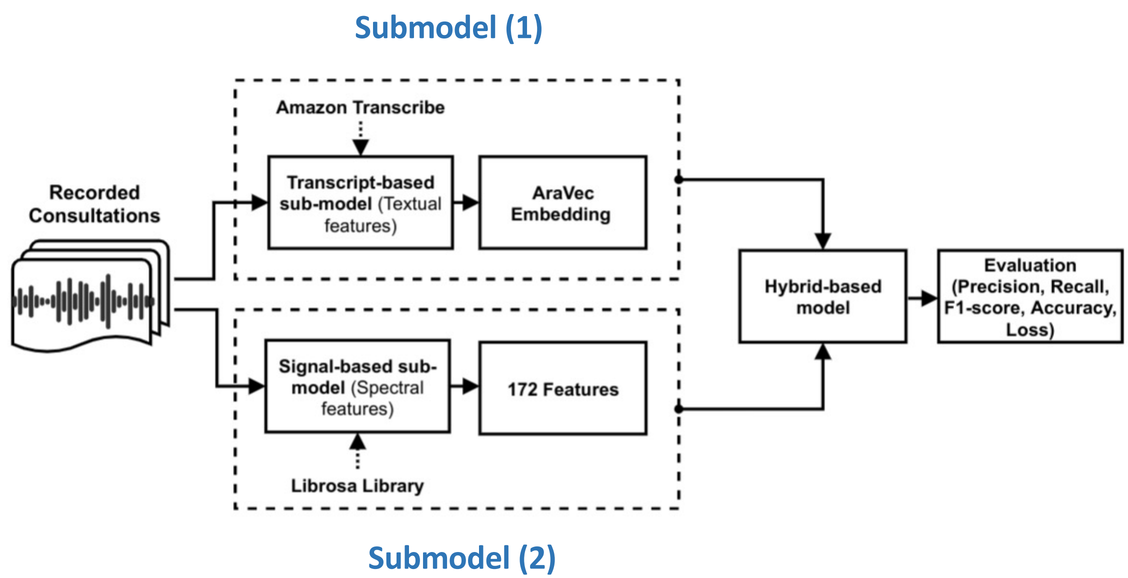
But **randomly sampling consultations** to find out **low-quality** one is not an efficient method and requires the operation team to listen to **a large number of consultations.**

For example. if we have a random sample of **100 consultations**, and the percentage of **poor**-consultations is **20%**; the operations team has to **check 100** consultations **to find 20 poor**-items in the best case.

The **objective** is to improve the precision of identifying the low-quality calls.

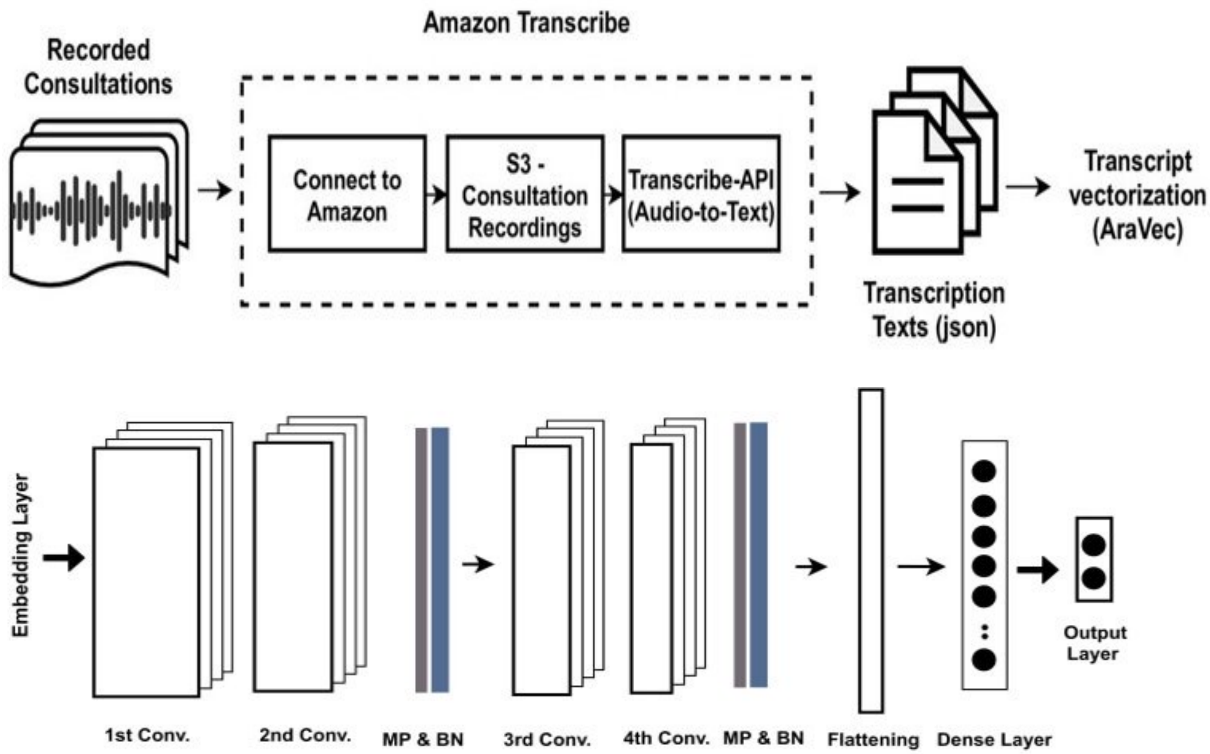
Quality Control
of Medical Consultations
over Voice

(Proposed Model)



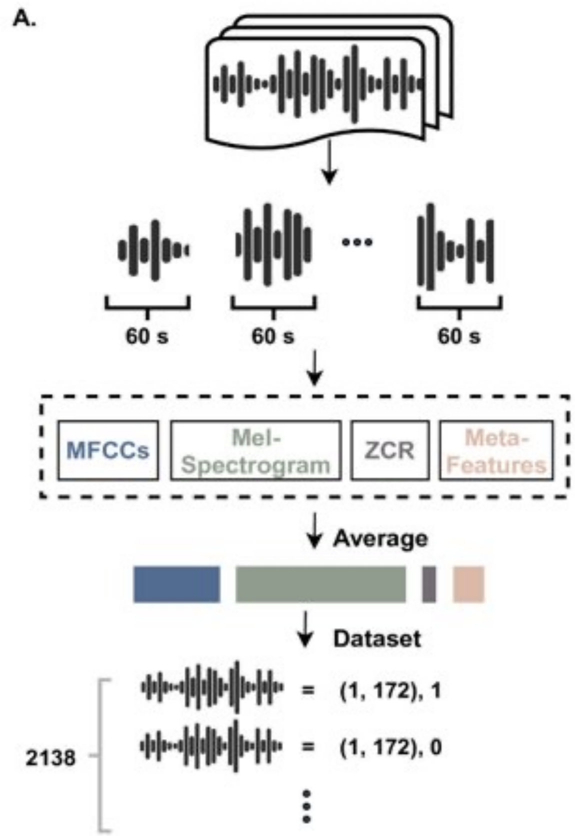
Quality
Control of Medical
Consultations over
Voice

(Text-based Module)



Quality
Control of Medical
Consultations
over Voice

(Signal-based features)



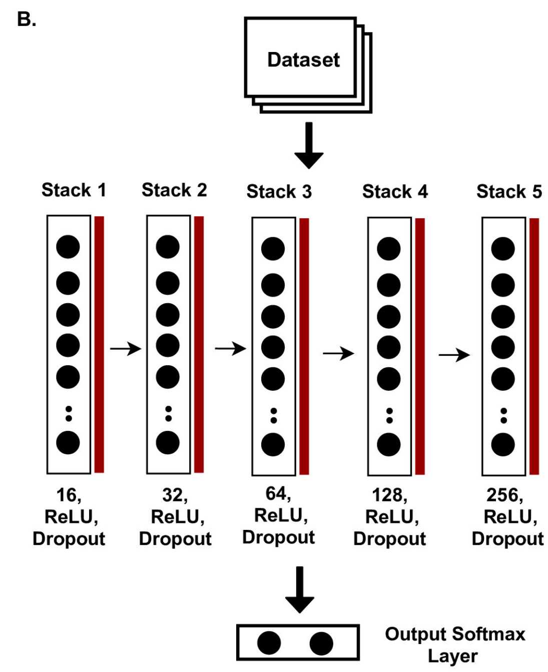
Samples per audio
= sampling rate ***** duration

No. of segments
= duration **/** segment length

Samples per segment
= samples per audio **/** No. of segments

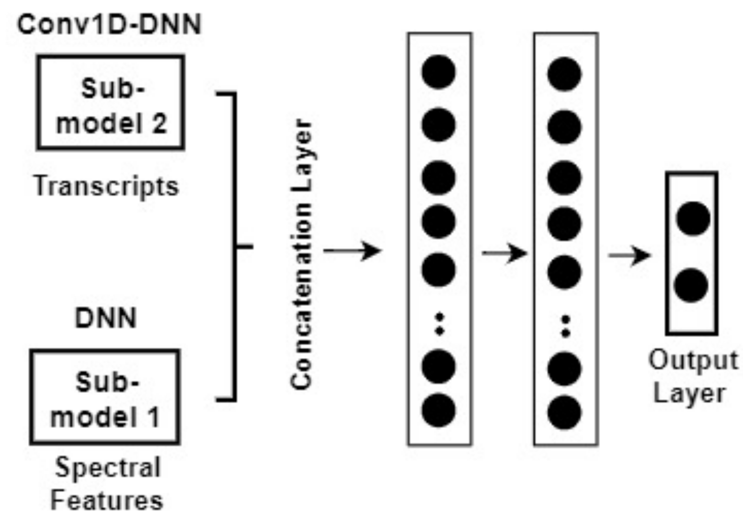
Quality Control of Medical Consultations over Voice

(Signals-based Model)



Quality Control of Medical Consultations over Voice

(Hybrid Model)



Quality Control of Medical Consultations over Voice (Results)

Results of the hybrid of the transcripts & signals-based submodels

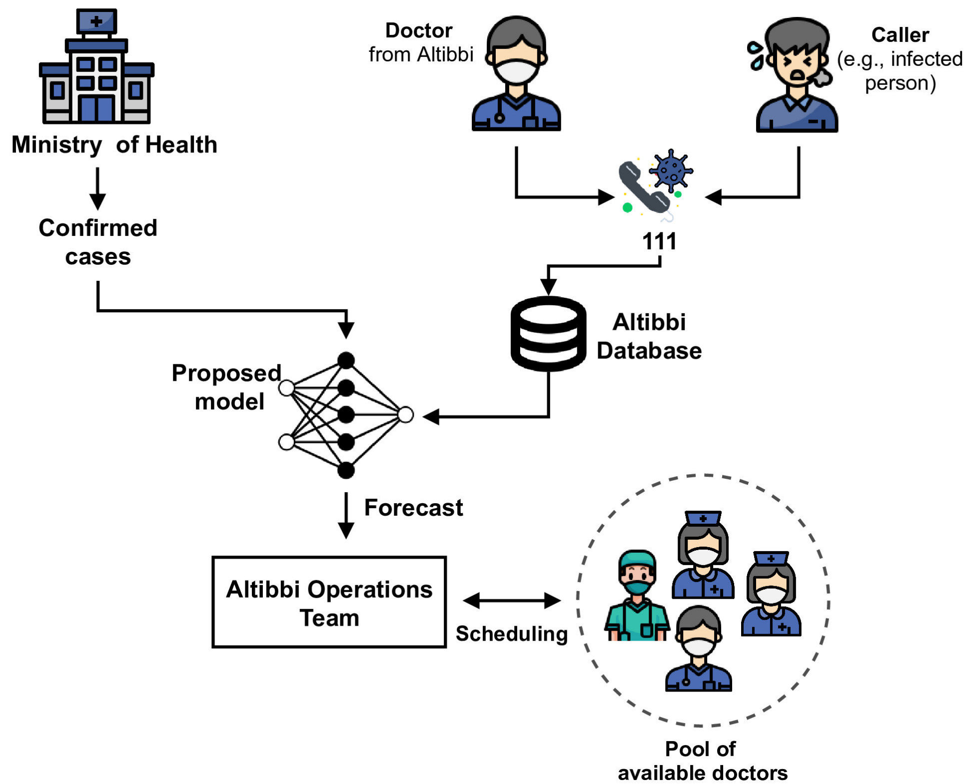
Table 5. The results of the hybrid approach of the transcripts and the spectral features using AraVec-Twitter at different structures of embedding models (E.M.), different E.D., embedding weights (E.W.), LR., and the batch size (B.S.).

Vocab Size	E.M.	Precision		Recall		F1-Score		Acc.	Loss	Epochs	E.W.	E.D.	LR.	B.S.
		P.C.	Mc. Avg.	P.C.	Mc. Avg.	P.C.	Mc. Avg.							
	SG	0.000	0.299	0.000	0.495	0.000	0.373	0.595	2.085				5E-04	
9000	SG	0.399	0.199	**1.000**	0.500	**0.570**	0.285	0.399	11.683	30	Non	300	9E-04	128
	SG	0.399	0.199	**1.000**	0.500	**0.570**	0.285	0.399	2.299				5E-05	
	SG	0.399	0.199	**1.000**	0.500	**0.570**	0.285	0.399	2.020				5E-04	
All	SG	0.000	0.301	0.000	0.500	0.000	0.375	**0.601**	11.685	30	Non	300	9E-04	128
	SG	0.394	0.322	0.977	0.491	0.562	0.286	0.393	2.162				5E-05	
	CBOW	0.400	0.501	0.250	0.501	0.308	0.488	0.551	3.315				5E-04	
All	CBOW	0.000	0.301	0.000	0.500	0.000	0.375	**0.601**	11.616	30	Non	100	9E-04	128
	CBOW	0.383	0.387	0.891	0.469	0.535	0.309	0.383	2.452				5E-05	
	CBOW	0.377	0.483	0.336	0.484	0.355	0.483	0.514	1.908				5E-04	
All	CBOW	0.408	0.510	0.586	0.511	0.481	0.495	0.495	11.623	30	Non	100	9E-04	64
	CBOW	**0.411**	0.512	0.523	**0.513**	0.460	**0.507**	0.511	1.524				5E-05	
	CBOW	0.397	0.489	0.898	0.496	0.550	0.355	0.414	1.256				5E-04	
All	CBOW	0.404	**0.513**	0.820	0.509	0.541	0.42	0.445	11.636	30	Non	100	9E-04	32
	CBOW	0.399	0.499	0.844	0.500	0.541	0.394	0.430	1.205				5E-05	
	CBOW	0.368	0.465	0.523	0.464	0.432	0.451	0.452	2.818	30				
All	CBOW	0.368	0.477	0.305	0.479	0.333	0.475	0.514	**0.932**	50	Trainable	100	5E-05	128
	CBOW	0.333	0.449	0.297	0.452	0.314	0.450	0.483	1.447	100				

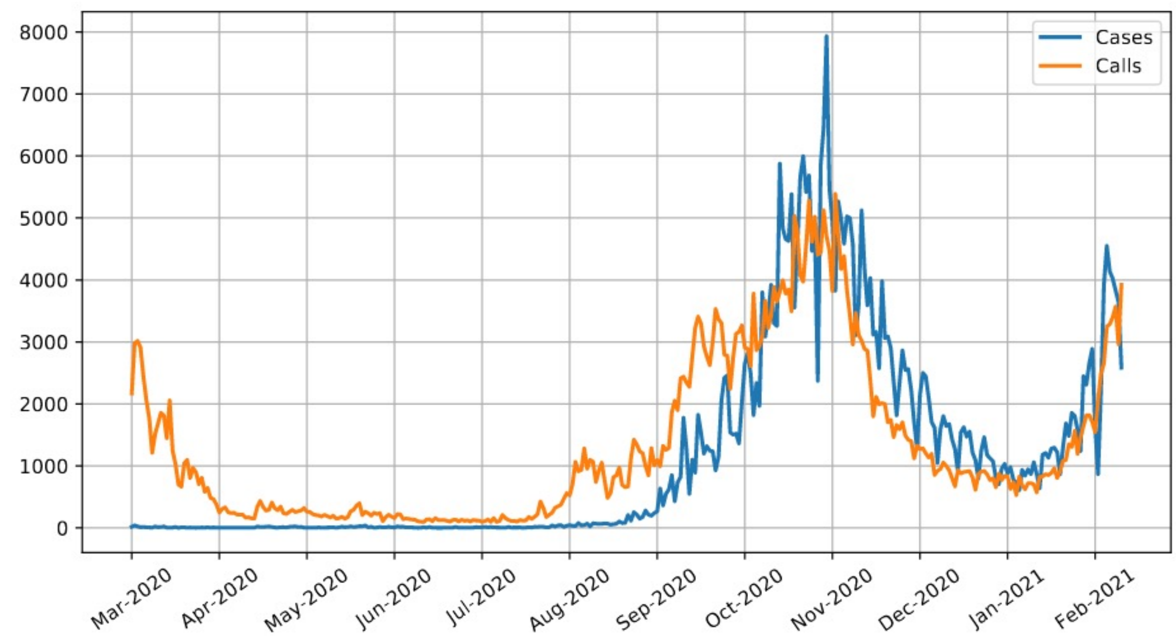
Quality Control of Medical Consultations over Voice (Challenges)

- The labeled dataset is relatively small, where a larger dataset might improve the model's performance.
- The data recorded at a low sampling rate (8khz), while 16khz is more recommended to capture various spectral and statistical features.
- Building a transcriber model for a speech of a mixture of dialects is very challenging.
- The definition of low-quality consultation is very broad.

Forecasting the number of received calls on COVID-19 hotline



Forecasting the number of received calls on COVID-19 hotline (Data)

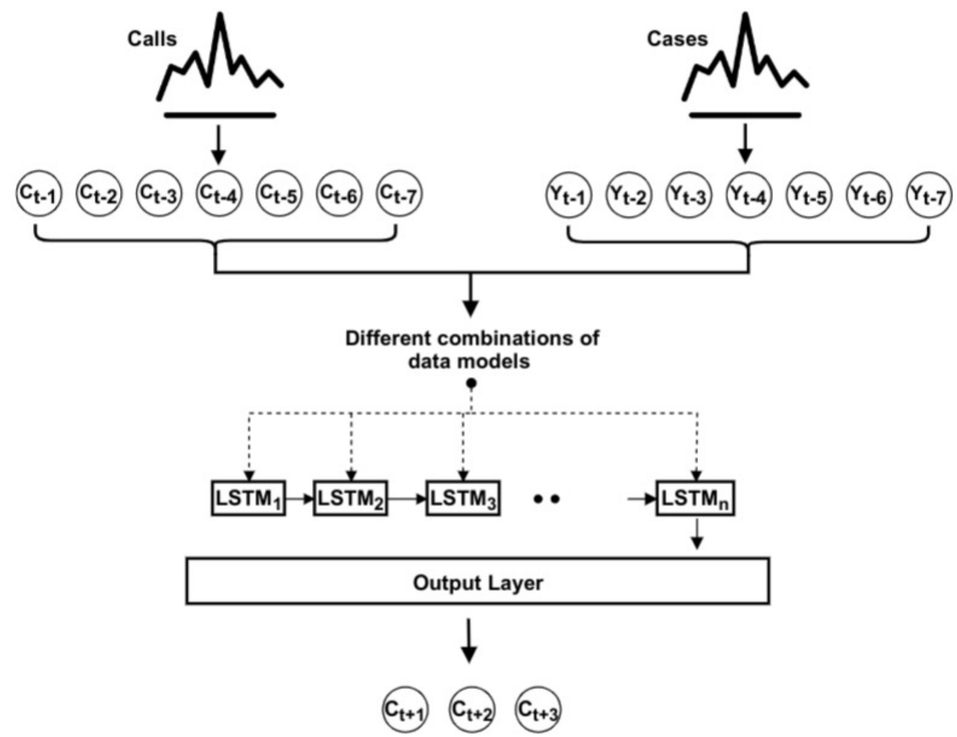


Forecasting COVID-19 calls (Models)

Table 1: Data Models for predicting the next three consecutive forecasts presented in terms of $C_{(t+1)}, C_{(t+2)}, C_{(t+3)}$.

Name	Model
$\mathbf{M_1}$	$f(C_{(t-1)})$
$\mathbf{M_2}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)})$
$\mathbf{M_3}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)})$
$\mathbf{M_4}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)})$
$\mathbf{M_5}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)})$
$\mathbf{M_6}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)}, C_{(t-6)})$
$\mathbf{M_7}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)}, C_{(t-6)}, C_{(t-7)})$
$\mathbf{M_8}$	$\mathrm{f}(C_{(t-1)}, Y_{(t-1)})$
$\mathbf{M_9}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, Y_{(t-1)}, Y_{(t-2)})$
$\mathbf{M_{10}}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, Y_{(t-1)}, Y_{(t-2)}, Y_{(t-3)})$
$\mathbf{M_{11}}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, Y_{(t-1)}, Y_{(t-2)}, Y_{(t-3)}, Y_{(t-4)})$
$\mathbf{M_{12}}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)}, Y_{(t-1)}, Y_{(t-2)}, Y_{(t-3)}, Y_{(t-4)}, Y_{(t-5)})$
$\mathbf{M_{13}}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)}, C_{(t-6)}, Y_{(t-1)}, Y_{(t-2)}, Y_{(t-3)}, Y_{(t-4)}, Y_{(t-5)}, Y_{(t-6)})$
$\mathbf{M_{14}}$	$\mathrm{f}(C_{(t-1)}, C_{(t-2)}, C_{(t-3)}, C_{(t-4)}, C_{(t-5)}, C_{(t-6)}, C_{(t-7)},$ $Y_{(t-1)}, Y_{(t-2)}, Y_{(t-3)}, Y_{(t-4)}, Y_{(t-5)}, Y_{(t-6)}, Y_{(t-7)})$

Forecasting COVID-19 calls (Models)



Forecasting COVID-19 calls (Results)

Table 4: The performance of LSTM based on the best tuned parameters, and the 14 constructed data models.

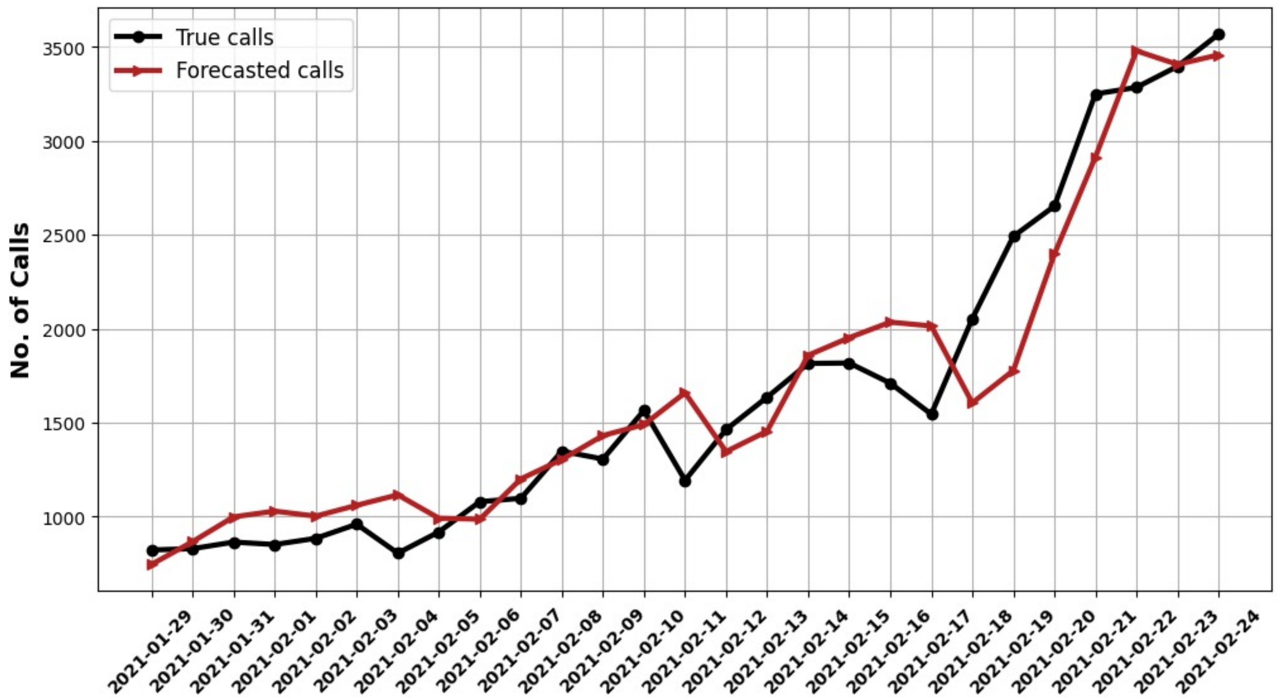
Model	$C_{(t+1)}$			$C_{(t+2)}$			$C_{(t+3)}$		
	RMSE	**MAE**	**R**	**RMSE**	**MAE**	**R**	**RMSE**	**MAE**	**R**
M_1	263.71	202.02	0.972	398.53	299.59	**0.936**	535.83	407.07	**0.919**
M_2	310.37	240.96	0.967	447.42	340.89	0.924	600.57	456.40	0.908
M_3	370.81	289.95	0.961	537.05	408.62	0.906	654.43	499.16	0.912
M_4	436.96	342.57	0.958	577.16	441.71	0.911	696.11	548.46	0.915
M_5	435.10	347.99	0.963	565.15	435.84	0.914	721.92	576.03	0.916
M_6	539.85	429.45	0.956	591.13	464.53	0.909	704.01	570.66	0.914
M_7	346.15	280.64	0.956	573.61	466.26	0.897	618.45	499.79	0.889
M_8	**244.56**	**184.57**	0.972	**367.72**	**274.06**	0.930	**511.05**	**368.01**	0.907
M_9	285.04	219.02	0.967	425.05	314.82	0.919	571.91	430.38	0.908
M_{10}	301.41	238.52	0.965	484.40	369.87	0.915	644.61	486.23	0.908
M_{11}	292.12	238.35	**0.974**	549.46	441.69	0.921	531.66	410.88	0.910
M_{12}	386.72	313.25	0.968	565.93	443.03	0.918	794.92	656.45	0.909
M_{13}	300.87	236.95	0.946	557.31	432.94	0.875	587.60	438.60	0.861
M_{14}	304.39	240.45	0.935	506.96	386.94	0.865	644.16	505.25	0.859

Forecasting COVID-19 calls (Results)

Table 5: The performance of BiLSTM based on the best tuned parameters, and the 14 constructed data models.

Model	$C_{(t+1)}$			$C_{(t+2)}$			$C_{(t+3)}$		
	RMSE	**MAE**	**R**	**RMSE**	**MAE**	**R**	**RMSE**	**MAE**	**R**
M_1	236.02	178.54	0.973	361.30	277.39	0.939	490.10	366.93	0.923
M_2	260.56	204.89	0.972	406.03	308.61	0.932	547.44	415.27	0.917
M_3	284.56	229.92	0.971	443.12	340.76	0.926	587.06	450.39	0.916
M_4	312.38	255.65	0.969	457.55	357.94	0.924	599.37	467.66	0.913
M_5	293.34	237.29	0.969	439.49	344.85	0.920	584.23	455.67	0.905
M_6	270.03	207.68	0.963	446.38	347.08	0.906	590.04	466.60	0.873
M_7	288.92	240.82	0.964	467.44	373.90	0.898	570.54	457.31	0.887
M_8	219.33	**164.64**	0.971	**347.05**	**265.29**	0.937	**479.49**	**348.65**	0.918
M_9	236.06	183.20	0.974	392.54	299.78	0.932	532.61	410.35	0.917
M_{10}	246.70	192.80	0.975	365.66	282.90	0.938	519.34	398.54	**0.924**
M_{11}	229.46	179.58	**0.976**	408.90	302.67	**0.943**	502.17	371.27	0.913
M_{12}	403.41	335.26	0.974	658.86	562.81	0.891	765.91	662.67	0.919
M_{13}	**207.76**	164.98	**0.976**	404.58	302.29	0.904	502.67	379.14	0.888
M_{14}	405.33	350.01	0.969	477.79	370.13	0.885	627.84	531.41	0.869

Forecasting COVID-19 calls (Results)



Forecasting COVID-19 calls (Challenges)

- Including other factors?
- Increasing the prediction horizon.

General Challenges

Related to the Language:

- Arabic has a very complex morphology.
- The use of mixture of dialects

Other challenges:

- The lack of in-domain datasets.
- Sharing the data and privacy.
- Interpretability of the models.
- Sensitivity of medical scenarios

List of Publications

- Faris, H., Habib, M., Faris, M., Alomari, M. and Alomari, A., 2020. Medical speciality classification system based on binary particle swarms and ensemble of one vs. rest support vector machines. Journal of biomedical informatics, 109, p.103525. Elsevier
- Faris, H., Habib, M., Faris, M., Alomari, A., Castillo, P.A. and Alomari, M., 2021. Classification of Arabic healthcare questions based on word embeddings learned from massive consultations: a deep learning approach. Journal of Ambient Intelligence and Humanized Computing, pp.1-17. Spriger
- Faris, H., Habib, M., Faris, M., Elayan, H., & Alomari, A. (2021). An intelligent multimodal medical diagnosis system based on patients' medical questions and structured symptoms for telemedicine. *Informatics in Medicine Unlocked*, *23*, 100513. Elsevier

Thanks
Any questions?

hossam.fari@ju.edu.jo