

Deep multitask ensemble classification of emergency medical call incidents combining multimodal data improves emergency medical dispatch

SUPPLEMENTARY MATERIAL

**Pablo Ferri¹, Carlos Sáez¹, Antonio Félix-De castro², Javier Juan-Albarracín¹, Vicent Blanes-Selva¹,
Purificación Sánchez-Cuesta² and Juan M García-Gómez¹**

¹Biomedical Data Science Lab (BDSLab), Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València (UPV), Valencia, Spain

²Conselleria de Sanitat Universal i Salut Pública, Generalitat Valenciana (GVA), Valencia, Spain

Corresponding Author: pabferb2@upv.es

APPENDIX 1

During-call data

Demographics data are defined by age, gender and risk group variables. Age is a numerical discrete feature, gender is a categorical binary variable (*male, female*) and risk group is a categorical multiclass variable, i.e., with multiple possible values, such as *asthmatic, allergic, cardiac, diabetic, neoplastic*, etc.

Circumstantial factors data are integrated by date and caller type variables. Regarding to this last one feature, it is a categorical multiclass variable, keeping information about the person or institution which effectuated the emergency medical call and taking values such as *police, red cross, the patient, a relative*, etc.

Clinical variables data are conformed by features providing relevant medical information. They are collected in a sequential manner during the call, registering a subset of them, from the total 71 variables available. A full list including all these variables is presented in Table 1. As can be observed in this table, clinical variables are categorical, presenting one possible value or multiple ones. An example of how clinical variables are registered during an emergency medical call could be: *previous trauma, yes; hemorrhage, yes; bleeding site, rectal bleeding; consequences of the clinic, severe blood loss*. In this example, four clinical variables were registered, with their associated values.

Table 1a. Clinical variables with some of their possible values. Certain variables have just one possible associated value, while others may exhibit multiple values. To ease presentation, possible values are limited to three in this table.

Variable	Possible values
Active arrhythmia	yes
Active suicide attempt	yes
Acute decompensation of mental illness	yes
Administration of medication	yes
Age	less than 1 year, over 70 years
Altered behavior	abnormal behavior, aggressiveness/agitation
Arterial vascular clinic	yes
Bleeding site	epistaxis, hematuria, melena
Blood glucose	abnormal
Blood or mucus in stool	no, yes
Breathing	absent, labored
Burn	yes
Causation of intake	autolysis attempt , medication error
Choking	yes

Table 1b. Clinical variables with some of their possible values. Certain variables have just one possible associated value, while others may exhibit multiple values. To ease presentation, possible values are limited to three in this table.

Variable	Possible values
Clinic start	abrupt, progressive
Clinic triggers	upsetting
Clinical evolution	stable without worsening
Consequences of the clinic	mild blood loss, moderate blood loss, severe blood loss
Constipation	yes
Consumption of toxic substances	yes
Cyanosis	yes
Death	yes
Diarrhea	yes
Dizziness	yes
Drug intake	no, yes
Dyspnoea	no, yes
Dysuria and / or hematuria	yes
Eating / bilious vomiting	yes
Epidemiological criteria	contact with contaminated samples, contact with diagnosed cases
Epidemiological infectious disease	yes
Existence of neurological focality	yes
Fever	over 38, over 39
Flu syndrome	yes
Gastrointestinal symptoms	yes
Hemorrhage	no, yes
Hypertensive crisis	yes
ICTUS code criteria	no
Impaired consciousness	yes
Impaired consciousness level	yes
Incident location	highway, inter-urban road, lakes or rivers and other inland waters
Injury severity	major, minor, moderate
Intake household product	yes
Intake of substance (medicine or toxic)	yes
Itchiness	yes
Medical history	cardiac pathology, copd, diabetes
Menstruation	yes
Nasal congestion	no, yes

Table 1c. Clinical variables with some of their possible values. Certain variables have just one possible associated value, while others may exhibit multiple values. To ease presentation, possible values are limited to three in this table.

Variable	Possible values
Number of injured	from 1 to 3, over 3
Ongoing birth	yes
Pain	abdomen, generalized, head, lumbar area
Pregnant	no, yes
Previous trauma	no, yes
Prior care	no, yes
Recovered unconscious	yes
Regular medication	impossible to obtain, insulin, oral antidiabetics
Relationship and contact level	absent, present
Seizures	yes
Sickness	yes
Signs of severity	no, yes
Skin alteration type	edema/swelling
Skin disorders	yes
Symptoms of glottic edema	yes
Time of evolution	over 24 hours
Toxic substance	heroin
Treatment	prescribed treatment for the clinical picture, psychiatric medication
Type of accident	aggression, collision, drowning
Unconscious	no, yes
Vegetative picture	no, yes
Venous vascular clinic	yes
Vomiting	yes
Without further information	yes

Finally, free text dispatcher observations are constituted by short sentences, written during the call and providing additional relevant information which cannot be recorded in a structured manner. Examples of free text dispatcher observations are *according to the caller epileptic crisis, he has drunk and taken pills, he is half-conscious with half-closed eyes* or *patient bleeds abundantly from the head after falling at home, they have just found it in a pool of blood.*

After-call data

After-call data is recorded at a time after the call. They provide reliable information about the real patient state. These data are integrated by: posterior physician diagnosis, standardized by ICD-9 codes, such as *syncope* (ICD 780.2) or *acute myocardial infarction* (ICD 410); maneuvers and procedures indicating if the patient was *intubated*, *reanimated*, *sedated*, *received surgery*, etc.; and hospitalizations and urgency stays with information about the department where the patient was treated, the amount of time he stayed there and his discharge code.

APPENDIX 2

Data pre-processing

Age, structured stationary discrete ordinal variable, was mapped to a fuzzy [1] representation through piecewise linear functions. [2] These membership functions, presented in Figure 1, were validated by physicians of the Health Services Department of the Valencian Community. This smoothing transformation was carried out to avoid sharp transitions derived from grouping in a small set of categories discrete ordinal variables with high cardinality in their values.

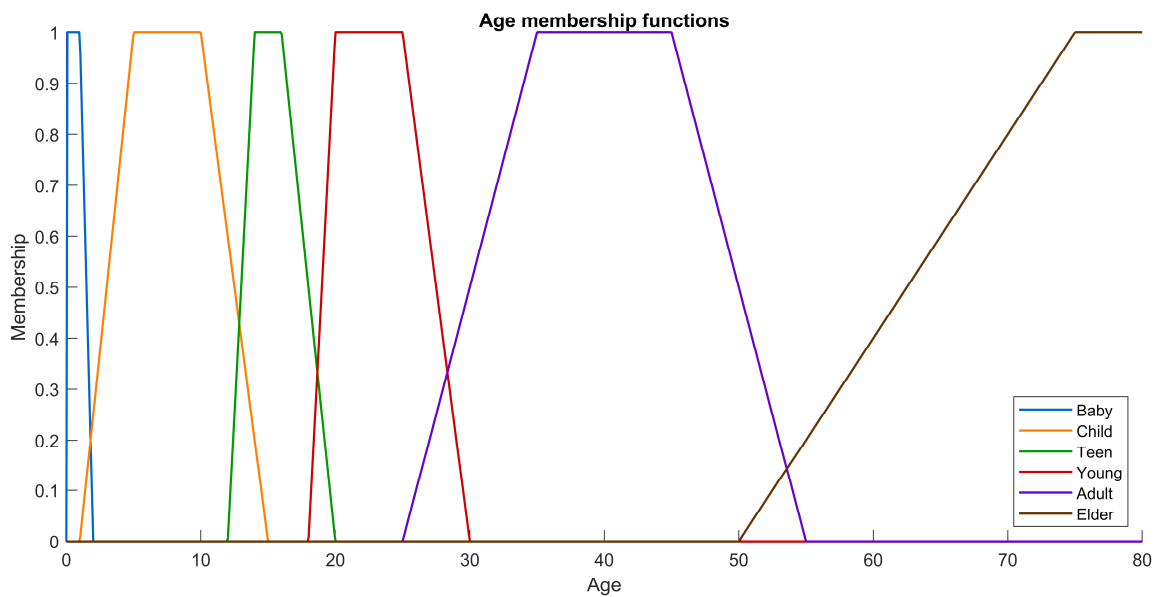


Figure 1. Piecewise linear functions representing age group membership.

Gender, risk group and caller type, structured stationary categorical variables, were one-hot encoded while several variables were derived from the date variable: weekday, month, if the day was or not a weekend day and if the day was or not was a bank holiday. These resulting variables, also structured stationary categorical variables, were one-hot encoded too.

Regarding to the clinical variables, structured sequential variables, each variable-value pair was converted to an integer, conforming then, sequences of integers that were pre-padded afterwards, to ensure sequences of fixed length. [3] This length was equal to 7, since in more than 99% of the incidents reported, the number of clinical variables collected was equal or lower than 7.

Likewise, spelling correction processes by means of fuzzy string matching [4] were applied over free text dispatcher observations, unstructured sequential variables, to reduce vocabulary dimensionality and noise. Besides, also to reduce vocabulary size, sub-word tokenization with WordPiece was carried out. [5] To ensure sequences of fixed length while keeping information about the original sequences lengths, post-padding and

attention mask generation were conducted. The padding length was set in 68, since in more than 99% of the incidents reported, the number of subwords written was equal or lower than 68.

Finally, labels, structured stationary categorical data, were one-hot encoded, deriving in a label matrix of 8 columns, each one associated with a specific label-class pair.

APPENDIX 3

Deep neural network design

The *Context subnetwork* consists on a multi-layer perceptron because this type of artificial neural network is suitable to model structured and stationary data. Fully connected layers in conjunction to leaky ReLU activation functions make possible to model potential non-linear relations among variables. Specifically, leaky ReLU functions are considered to avoid vanishing and exploding gradients, while preventing dead neurons issues. [6] Through batch normalization, covariate shift problems are managed, [7] while dropout eases regularization during training preventing neuron co-adaptation. [8] Likewise, softmax functions are used in the output blocks to dispose of a normalization score (between 0 and 1) for each class of each predicted label.

The *Clinical subnetwork* consists on a recurrent model, since clinical variables are notified in a sequential manner, being this recording order potentially informative. Long-short term memory (LSTM) units are used due to its ability to capture long-term dependences, far better than standard recurrent neural networks models. The usage of skip connections, as well as the fact that LSTM units are bidirectional, reduces the risk of losing relevant information during LSTM propagation. The final multi-layer perceptron module acts as an intermediary between the LSTM outputs and the final label predictions.

The *Text subnetwork* is mainly composed by a bidirectional encoding representations from transformers (BERT) block, since this model is at the state of the art in natural language processing tasks, [9] including text classification. As in the prior network, the multi-layer perceptron module eases to relate BERT outputs with output labels.

Finally, the *Ensemble subnetwork* is a multi-layer perceptron model, similar to the *Context subnetwork*, as this artificial neural network architecture is appropriate for its input data. It takes inner outputs of the prior subnetworks, not the final output scores, since dimensionality and processing level of these inner outputs present higher modeling potential than those relative to the final output scores of each network, which aggregate lots of information in a small set of scalar values.

APPENDIX 4

Hyperparameter tuning

Table 2 shows the hyperparameters studied in our work, related with subnetworks architecture and optimizer settings. Their *optimal* values, i.e., values obtained after carrying out the multi-step hyperparameter tuning process, are also presented, for each network:

Table 2. Subnetworks hyperparameters with its definitive (*optimal*) values.

Hyperparameters	Model			
	Context net	Clinical net	Text net	Ensemble net
Embedding dimension	-	8	96	-
DB hidden layers	1	1	-	1
DB neurons per layer	64	64	-	256
DB dropout	0.1	0.1	-	0.25
BLSTM layers	-	6	-	-
BLSTM neurons	-	64	-	-
BLSTM dropout	-	0	-	-
Encoder attention heads	-	-	3	-
Encoder layers	-	-	3	-
Encoder neurons per layer	-	-	96	-
Encoder dropout	-	-	0.1	-
TSB layers	2	2	2	2
TSB neurons	64	64	64	64
Batch size	128	64	64	128
Learning rate	0.0005	0.0005	0.0005	0.0001
Weigth decay	0.00005	0.00005	0.0005	0.0005

Abbreviations: DB, dense blocks; TSB, task-specific blocks.

REFERENCES

- [1] L. A. Zadeh, "Fuzzy sets," *Information and control*, pp. 338-353 (8), 1965.
- [2] V. Novák, I. Perfilieva and J. Mockor, "Mathematical principles of fuzzy logic," *Springer Science & Business Media*, p. Vol. 517, 2012.
- [3] M. Dwarampudi and N. V. Reddy, "Effects of padding on LSTMs and CNNs," *arXiv preprint arXiv:1903.07288*, 2019.
- [4] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *Journal of the ACM (JACM)*, pp. 168-173 (21), 1974.
- [5] Y. Wu, M. Schuster, Z. Chen, Q. V. Le and M. Norouzi, "Google's neural machine translation system: Bridging the gap between human and machine translation.," *arXiv preprint arXiv:1609.08144*, 2016.
- [6] C. Nwankpa, W. Ijomah, A. Gachagan and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.
- [7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [8] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.
- [9] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding.," *arXiv preprint arXiv:1810.04805*, 2018.