Towards the intelligent diagnosis of hematological diseases

Sergio Díaz-del-Pino, Roberto Trelles-Martinez, Esteban Perez-Wohlfeil, Oswaldo Trelles

In traditional medicine, patient diagnosis usually implies an in depth study of its state and symptoms that a specialist has to carry out. The adaptation and customization of the medical treatment to those individual characteristics of each patient is what we know as Precision Medicine. However, the whole process requires several tests, analysis and administration procedures which in the end consumes time and makes it expensive for the public health system. Therefore, the research of new methods that assists the specialists could lead us to reduce not only the economic effect of the process but the workload while producing more accurate diagnosis and better quality of life of the patients.

Furthermore, in the case of multidisciplinary fields such as haematology, the identification of several diseases usually implies complex analyses such as peripheral blood smear, capillary electrophoresis and chromatography, in order to have a high degree of certainty in the diagnosis. A better understanding of the clinical tests and their relationship and the finding of new patterns between them will enable us to avoid a significant amount of such tests by supporting the specialist with new information.

In this line, Artificial Intelligence has proven to be a useful methodology for data analytics in general whose main drawback is the need of huge amounts of data to achieve high accuracy. In the particular case of clinical data, it is widely generated in hospitals but the lack of standardization and the difficulties of availability require complex preprocessing. Therefore, we have collected 100,000 complete blood counts (CBC) from the Hospital Clínico San Carlos (Madrid) and developed a method to 1) automatically label textual diagnosis using deep neural networks with Long short-term memory cells. In this approach, a group of specialists has manually labelled 1,000 CBCs through a mobile application, which have then been used to feed the network in order to learn to interpret the diagnosis, and 2) to make an intelligent diagnosis of new samples in which a subset of 10,000 CBCs has been used as an input to a Support Vector Machine.

In summary, in this work we present two different prototypes of architectures in order to define methods for the collection, preprocessing and intelligent classification of clinical data, focusing in haematological disease. Our proposal presents encouraging results with accuracies greater than 90% in both cases.