Automatic Sleep-Wake Stage Scoring using Artificial Neural Networks: Optimisation and Comparison of two Systems

Florian CHAPOTOT¹, Guillaume BECQ¹, Emilia SFORZA² et Alain BUGUET³

¹Département des Facteurs Humains, Centre de Recherches du Service de Santé des Armées, Grenoble, FRANCE.
²Laboratoire du Sommeil, Département de Psychiatrie, Hôpitaux universitaires de Genève, Genève, SUISSE.
³Institut de Médecine Tropicale, Service de Santé des Armées, Grenoble, FRANCE.

In human and animal, artificial neural networks have shown satisfying performance as compared to human experts in sleep-wake stage scoring analysis of polysomnographic (PSG) recordings (1-2). To date, the few available analysis systems cannot be used with any recordings because they have been developed using fixed parameters. These parameters consist of the number, type (EEG ± EMG ± EOG) and sampling rate of signals required, the recording device and digital file format used for data storage, as well as the number of sleep-wake stages to be determined.

A new software system for off-line polygraphic reviewing and analysis (PRANA, PhiTools, Strasbourg, France) has been developed under MATLAB (The MathWorks, Natick, USA). This environment, which supports virtually any recording systems, serves as a basis for developing new analysis and detection algorithms since it allows incorporating user software plug-ins. By doing so we have developed an automatic sleep-wake stage scoring analyser allowing to configure, learn and simulate artificial neural network classifiers. In its original version this automatic analyser software plug-in showed a global agreement of 74.7 ± 8.0% in scoring healthy human and rat recordings, respectively (4).

Objective: The aim of this study was to optimise the PRANA software automatic analyser plug-in for healthy human sleep-wake recordings and to compare its performance to that of another commercially available system (BioSleep v3.0 Oxford BioSignals, Oxford, Royaume-Uni). The parameters taken into account in the optimisation process were the use of a consensual multi-individual learning database, on one hand, and the introduction during simulation of a contextual dependency, on another hand.

Material & Methods: Nocturnal PSG recordings performed in 13 young adult healthy volunteers were interpreted independently by two experts (ES and AB) according to the standard R&K criteria (5), then submitted to automatic analysis by PRANA and BioSleep. The parameters taken into account in the optimisation process were the use of a consensual multi-individual learning database, on one hand, and the introduction during simulation of a contextual dependency, on another hand.

Results: Comparison of the two systems against the experts indicate better performances for the PRANA system than for BioSleep (global agreement of 79.5±6.3 versus 46.2±8.7%). Intra-class agreements and errors are represented in the accompanying figure.

Conclusion: A global performance of the PRANA system just below inter-expert agreement (82.8±3.3%) suggest the use of the system at the sleep lab.