**Problem statement:** Current biofluid diagnostic tests for common pathologies and monitoring tests like blood glucose levels in blood present limitations in their use and application. Tests like RT-PCR (real time polymerase chain reaction) for SARS-COV-2, for example, necessitate the use of precise biochemical labels. On top of this, it is a very limited test that makes it difficult to apply to a large scope of pathologies without having access to very advanced biochemical equipment and methods. The development of a new, rapid, non-invasive, and adaptable diagnostic test is necessary to address this problem. Raman spectroscopy presents itself as an ideal candidate as it is a low-cost, non-invasive, label-free, and versatile method enabling reliable detection of multiple pathologies. Recently, a SARS-COV-2 detection method using Raman spectroscopy on saliva samples has proven to be reliable and efficient in obtaining early diagnostics. This research, coupled with advances in biofluid preparation protocols have opened new outlooks on the applications of Raman spectroscopy into such fields as cholesterol, glucose, iron levels and thyroid hormones monitoring. Such efforts are currently hindered by the lack of predictive, machine learning models that can take account of confounding factors such as age, sex, hormone levels, etc.

**Objectives:** The main goal of this project is to develop a bank of machine learning classification models enabling the expansion of the application scope of Raman spectroscopy diagnostic imaging. The following sub-objectives have been identified to achieve this goal:

OBJ : Acquire and prep a biobank of more than 1000 biofluid samples and Raman spectra for processing

Obj: Stratify the acquired dataset in terms of demographics, standard molecular test results and other confounding factors

Obj: Produce multiple predictive models (machine learning) enabling compensation for confounding factors.

**Proposed Methodology:** In recent years, the Laboratory of Radiological Optics (LRO) has been developing biofluid collecting, prepping, and imaging protocols enabling accurate and uniform Raman spectra datasets. LRO has also been developing machine learning frameworks enabling quick and versatile classification modeling from these datasets. The planned acquisition of a massive biobank of samples and spectra along with expansive medical information on more than 1000 patients will create an opportunity to train multiple new classification models. These classification models will then be tested and optimized to achieve high sensibility and specificity.

**Importance of Research:** The realization of this project would lay the groundwork for numerous mobile, fast, low-cost biofluid diagnostic tests in the future. Saliva tests have previously been effective in diagnosing concussions. The classification model bank would permit the rapid transposition of the Raman spectroscopy diagnosis imaging method to this application. Other diagnostics such as cholesterol, iron levels and thyroid hormone levels could then be integrated using the same equipment and sample collection protocols. Such a device would be revolutionary in the rapid-diagnostic space with incomparable versatility and speed.

[1] K. Ember, F. Daoust, M. Mahfoud, F. Dallaire, E. Zamani, Salive-based detection of COVID-19 infection in a real-worls setting using reagent-free Raman spectroscopy and machine learning, *Science Advances,* (2021).