

## DETTAGLIO OBIETTIVI REALIZZATIVI

OR	TITOLO
OR 1	REALIZZAZIONE DI UN NETWORK DI BIOBANCHE PER BIOPSIE LIQUIDE E TISSUTALI PER LE PATOLOGIE DEL SISTEMA NERVOSO CENTRALE E VASCOLARE
<b>Att. 1.1</b>	Estensione della mappatura delle biobanche/banche dati dal contesto regionale campano alle sedi degli altri partner di progetto. Analisi dei criteri di inclusione
<b>Att. 1.2</b>	Progettazione e sviluppo di una infrastruttura fisica e virtuale per l'integrazione delle biobanche/banche dati
<b>Att. 1.3</b>	Messa a punto di metodologie e tecniche per l'estrazione e la correlazione dei dati
<b>Att. 1.4</b>	Progettazione e sperimentazione di modelli predittivi per le patologie neuro-psichiatriche e vascolari

ATTIVITA'	DESCRIZIONE
1.1	Estensione della mappatura delle biobanche/banche dati dal contesto regionale campano alle sedi degli altri partner di progetto. Analisi dei criteri di inclusione

L'attivazione dell'iter procedurale per l'accreditamento UNI ISO 20387:2019, non previsto nella fase di presentazione del progetto, rappresenta un requisito rilevante ai fini della corretta analisi dei criteri di inclusione e per la mappatura delle biobanche/banche dati afferenti alle diverse sedi dei partner di progetto.

Ciò premesso, le evoluzioni del contesto di riferimento e gli scenari futuri ipotizzabili in merito alla pandemia da COVID-19, hanno reso ancor più evidente la necessità di focalizzare le risorse sullo standard di accreditamento al fine di estrapolare i requisiti di riferimento per le Biobanche.

**I contenuti sono stati puntualmente valutati e trasposti in termini di requisiti da implementare e strutturare come asset portante del software di gestione della Biobanca – e del portale per l'integrazione delle biobanche – in modo tale da garantire un efficace governo dei processi e la presenza di un archivio dei documenti e dati progressivamente implementati basato su una logica strutturata e tracciabile.**

*Data la sospensione totale o parziale delle attività in ragione delle disposizioni restrittive emanate in conseguenza all'emergenza sanitaria COVID19, tali attività sono proseguite esclusivamente in modalità agile senza possibilità di ispezioni in loco.*

RI	SDN (BIOCAM) , NEATEC (BIOCAM) , UNISA, UNIMI, UNIBS, UNISANRAFFAELE, CCM	98%	Analisi della norma UNI ISO 20387 “Biotechnology – Biobanking – General requirements for biobanking” Deliverables 1.1.1 e 1.1.2 in revisione finale. Se ne prevede la consegna alla chiusura dell'attività
----	---	-----	--

ATTIVITA'	DESCRIZIONE
1.2	Progettazione e sviluppo di una infrastruttura fisica e virtuale per l'integrazione delle biobanche/banche dati.

È stato sostanzialmente **completato lo sviluppo del Portale** per l'integrazione delle biobanche progettato e descritto in precedenza. In particolare sono state realizzate le sezioni:

1. Home,
2. Biobanche,
3. Pubblicazioni,
4. Catalogo (con le funzionalità di ricerca semplice e avanzata),
5. News.

Sono stati anche effettuati dei **test funzionali e prestazionali** con dati di prova costruiti ad hoc, ossia delle **simulazioni dei dati reali** che dovrebbero essere forniti dalle Biobanche coinvolte nell'OR1.

*Purtroppo, la sospensione delle attività di laboratorio in ragione delle disposizioni restrittive emanate in conseguenza all'emergenza sanitaria COVID19, non ha permesso di eseguire test di funzionalità con gli utilizzatori finali di esso (es. ricercatori e operatori sanitari delle aziende/università coinvolte nel progetto).*

SS	NEATEC (BIOCAM) UNISANRAFFAELE	70%	Implementazione della base dati da utilizzare per contenere le informazioni gestite dal Portale. Deliverable 1.2.1 in draft
----	-----------------------------------	-----	--

ATTIVITA'	DESCRIZIONE
1.3	Messa a punto di metodologie e tecniche per l'estrazione e la correlazione dei dati.

È proseguita la sperimentazione per l'estrazione dati mediante tecniche di intelligenza artificiale dai file di testo non strutturato che rimane completamente inutilizzato nelle analisi clinico-statistiche tradizionali, pur essendo estremamente ricco di contenuti informativi legati all'oggetto di analisi.

Per superare questa contraddizione abbiamo deciso di ricorrere a tecniche di Natural Language Processing (NLP), che consentono di comprendere il testo e la semantica dello stesso, attraverso l'implementazione di una **pipeline semantica** atta alla comprensione e alla disambiguazione del testo, calata nel dominio clinico, composta, tra l'altro, dalla:

- a) Lemmatizzazione
- b) Tokenizzazione
- c) Eliminazione delle stopwords
- d) Identificazione delle entità tramite Named Entity Recognition.

Ai fini del potenziamento dell'information retrieval è stato effettuato il **word embeddings** che ibrida da una parte l'analisi morfologica, l'analisi sintattica del periodo, l'identificazione delle entità e il linking delle stesse e dall'altra il Reasoning Ontologico. Si unisce così il mondo GOFAI (Gold Old-Fashioned AI), il Machine Learning (classico) ed il Deep Learning.

RI	NEATEC (BIOCAM) UNISANRAFFAELE	45%	Valutazione sperimentale delle varie tecniche per l'estrazione e la normalizzazione dei dati da utilizzare negli algoritmi predittivi Deliverables 1.3.1 e 1-3-2 in fase di stesura
----	-----------------------------------	-----	--

ATTIVITA'	DESCRIZIONE
1.4	Progettazione e sperimentazione di modelli predittivi per le patologie neuro-psichiatriche e vascolari.

La classificazione sistematica delle “features” di pazienti con patologie di interesse neurologico, ha facilitato la costruzione di un dataset dedicato alla modellazione di un DSS clinico basato su tecniche di AI per la previsione del rischio di abuso farmacologico in pazienti con emicrania, conclusasi con l'accettazione per la pubblicazione da parte della rivista Computational & Structural Biotechnology Journal.

È, inoltre, proseguita l'analisi dell'associazione di alcune varianti geniche con le WMHs (white matter hyperintensities), un marcatore neuroradiologico associato ad aumentato rischio di ictus, demenza e morte. L'analisi dei dati ha mostrato che:

1. l'allele LTA rs2071590T (OR=2,2; p=0,034) e l'allele SOD rs2234694C (OR=4,9; p=0,003) sono entrambi associati alla presenza di WMH in modelli classici di trasmissione ereditaria.
2. la frequenza di un aplotipo a quattro loci (aplotipo TGCT) è significativamente più alta negli emicranici con WMH (18,4%) rispetto a quelli senza (3,0%)(OR=8.7, p=0.009).
3. uno stato di portatore di aplotipo TGCT è associato in maniera indipendente a un aumento del rischio di WMH alla risonanza magnetica

È in corso l'applicazione del modello di apprendimento automatico sopra descritto alla previsione del rischio di complicanze cerebrovascolari in questo subset di pazienti.

RI	NEATEC (BIOCAM) UNISANRAFFAELE	50%	Applicazione di un sistema decisionale clinico ad una popolazione di pazienti di interesse neurologico Deliverables previsti a fine progetto
----	-----------------------------------	-----	---

# PRODOTTI DELLA RICERCA

---

OR1	REALIZZAZIONE DI UN NETWORK DI BIOBANCHE PER BIOPSIE LIQUIDE E TISSUTALI PER LE PATOLOGIE DEL SISTEMA NERVOSO CENTRALE E VASCOLARE
1	Barbanti P., Aurilia C., Egeo G., Fofi L., Guadagni F., Ferroni P.: Dopaminergic symptoms in migraine: a cross-sectional study on 1148 consecutive headache center-based patients. <i>Cephalalgia</i> , 40(11):1168-1176, 2020
2	Ferroni P., Zanzotto F.M., Scarpato N., Spila A., Fofi L., Egeo G., Rullo A., Palmirota R., Barbanti P., Guadagni F.: Machine learning approach to predict medication overuse in migraine patients. <i>Comput. Struct. Biotechnol. J.</i> , 18:1487-1496, 2020
3	Ferroni P., Barbanti P., Aurilia C., Egeo G., Fofi L., Valente M.G., De Marchis M.L., Spila A., Palmirota R., Guadagni F.: MS SOD and White Matter Hyperintensities in migraine. (In Preparation), 2020