



UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA

Corso di Laurea triennale in Ingegneria Biomedica

**Analisi con tecniche di clustering di curve da
test di tolleranza al glucosio per via orale**

**Clustering analysis of curves from oral glucose
tolerance tests**

Relatore:

Dott.ssa Micaela Morettini

Correlatori:

Dott. Andrea Tura

Rapporto Finale di:

Martina Fratangelo

Anno Accademico 2021/2022

Abstract

Il diabete è una patologia estremamente diffusa al giorno d'oggi in tutto il mondo. Sono infatti sempre di più i soggetti che sviluppano il diabete (soprattutto il diabete di tipo 2) o che si trovano comunque in una fase di prediabete, con possibilità di sviluppo verso la patologia conclamata. Lo scopo del presente studio è quello di analizzare con tecniche di clustering la forma delle curve di glucosio (da 0 a 120 minuti), insulina (da 0 a 120 minuti) e Peptide-C (da 0 a 180 minuti) ottenute da test di tolleranza al glucosio per via orale nell'ottica di ottenere sottogruppi di soggetti e valutarne le differenze in base alle loro caratteristiche fisiopatologiche. Nello specifico, è stato utilizzato l'algoritmo "k-means", secondo tre metodologie differenti (denominate "Calinski", "Silhouette" e "Davies"), per definire la soluzione ottima di cluster per l'intera popolazione a partire dalle curve di glucosio, insulina, glucosio ed insulina insieme, Peptide-C, e tutte e tre le curve insieme. Definito il numero di cluster, questi sono stati poi confrontati valutando delle caratteristiche scelte per ogni soggetto così da poter definire le differenze fra un sottogruppo e l'altro. L'analisi ha portato alla definizione di 2 o 3 cluster diversi fra loro: infatti, nonostante categorie di soggetti con diverso stato di tolleranza al glucosio (normale tolleranza al glucosio, NGT, alterata tolleranza al glucosio o alterata glicemia a digiuno IGT/IFG, diabete mellito di tipo 2, T2DM) fossero presenti in numero variabile in tutti i cluster (con prevalenza di NGT nel cluster 1 e di IGT/IFG e T2DM nel cluster 2), i risultati hanno evidenziato che le diverse "shape" delle curve temporali, attribuite ai vari cluster e rappresentanti prevalentemente soggetti NGT, possono caratterizzare anche soggetti T2DM. Viceversa, soggetti NGT erano presenti nei cluster maggiormente associati a condizioni di IGT/IFG e T2DM. In conclusione, lo studio è quindi stato utile per evidenziare che un clustering basato sulla differente forma delle curve temporali di glucosio, insulina e Peptide-C può rilevare la presenza eterogenea di soggetti NGT, IGT/IFG e T2DM, e potrebbe dare informazioni aggiuntive rispetto alla classificazione basata sui criteri diagnostici standard.

INDICE

I. Introduzione.....	4
1. Omeostasi del glucosio e sue alterazioni: il diabete	
1.1 Omeostasi del glucosio.....	5
1.2 Fisiopatologia.....	7
1.3 Diagnosi.....	8
1.4 Classificazione.....	11
1.4.1 Diabete di tipo 1.....	11
1.4.2 Diabete di tipo 2.....	12
1.4.3 LADA.....	12
1.4.4 MODY.....	12
1.4.5 Diabete gestazionale.....	13
2. Analisi con tecniche di clustering di curve da test di tolleranza al glucosio per via orale	
2.1 Stato dell'arte.....	14
2.1.1 Clustering basato su OGTT proposto da Wagner et al.....	14
2.1.2 Clustering basato su OGTT proposto da Obura et al.....	17
2.2 Definizione popolazione di studio.....	19
2.3 Clustering.....	19
2.4 Variabili selezionate.....	21
2.5 Analisi Statistica.....	22
2.6 Risultati.....	23
II. Discussione e Conclusione.....	37

III. Bibliografia.....38

I. Introduzione

Il diabete mellito rappresenta una patologia estremamente eterogenea basata sulla perdita progressiva della secrezione insulinica (diabete di tipo 2), conseguenza dell'insulino-resistenza, o sulla completa autodistruzione delle cellule β (diabete di tipo 1), con successiva carenza assoluta dell'ormone nell'organismo. Nel 2021 l'International Diabetes Federation (IDF) ha calcolato che oltre 530 milioni di persone nel mondo tra i 20 e i 79 anni sono diabetiche. In particolare, gli italiani affetti da diabete di tipo 2 sono circa il 6% della popolazione, per un totale di circa 4 milioni di individui. A coloro che hanno già sviluppato la patologia (DM), si aggiungono i soggetti in fase di prediabete (Impaired Glucose Tolerance, IGT e/o Impaired Fasting Glucose, IFG) che presentano alterati valori della glicemia, che non rientrano tuttavia nelle condizioni di diabete. Inoltre, solo circa il 50% dei soggetti con regolazione glicemica compromessa (quindi in fase di prediabete) progredisce verso la patologia conclamata. Pertanto, la determinazione dei sottotipi del prediabete e del diabete diventa essenziale per la prevenzione personalizzata della patologia. Per questo motivo, la distinzione fra un soggetto sano, uno in prediabete (IFG e/o IGT) ed uno affetto da diabete di tipo 2 è sempre stata oggetto di studio da parte della ricerca, presentandosi ancora tutt'oggi come una delle più grandi sfide nella diagnosi precoce.

Per giungere alla soluzione di questo problema, gli studiosi si sono concentrati fin'ora sull'utilizzo di tecniche di Machine Learning e, nello specifico, di algoritmi di clustering, come il "k-means": hanno confrontato, per ogni paziente da esaminare, i valori delle features caratteristiche della patologia (come glicemia a digiuno, sensibilità e secrezione dell'insulina, ecc..), ottenute in maggioranza attraverso OGTT di durata variabile, con quelli soglia tipici di ogni condizione, così da definire lo status di salute di ogni soggetto.

Il presente studio si pone lo scopo di classificare i pazienti utilizzando sempre algoritmi di clustering, partendo, però, non dalle variabili ma dalle curve temporali di glucosio (da 0 a 120 minuti), di insulina (da 0 a 120 minuti) e di Peptide-C (da 0 a 180 minuti), così da poter definire, in base alla forma della curva di un determinato soggetto, il relativo status di salute e la relativa appartenenza di questo ad un dato sottogruppo (NGT, IGT/IFG, T2DM).

1. OMEOSTASI DEL GLUCOSIO E SUE

ALTERAZIONI: IL DIABETE

Il diabete mellito non può essere considerato come una singola malattia, ma piuttosto come un gruppo eterogeneo di sindromi che abbraccia tutti gli aspetti metabolici dell'organismo (in particolare quelli del glucosio) e che, con le sue complicanze, interessa tutti gli organi e gli apparati di un essere vivente. È una condizione morbosa diffusa in tutto il mondo con un'incidenza variabile nelle differenti popolazioni in rapporto alle condizioni socio-economiche che si riflettono sul tenore alimentare, in particolare sulla qualità della dieta e sull'attività fisica. Secondo i dati Istat, oltre 3 milioni 200 mila persone in Italia dichiarano di essere affette da diabete (il 5,3% dell'intera popolazione, superando il 16% fra le persone con oltre 65 anni), con una diffusione che è quasi raddoppiata negli ultimi trent'anni, in ragione, tra l'altro, dell'invecchiamento della popolazione e di più precoci diagnosi e migliori cure (le quali aumentano la sopravvivenza dei pazienti affetti da questa patologia) [1-2-3].

1.1 OMEOSTASI DEL GLUCOSIO

Negli esseri viventi esistono specifici minerali e molecole che svolgono funzioni vitali per l'economia dell'organismo. È pertanto necessario mantenere costanti i loro livelli ematici, al fine di evitarne improvvise oscillazioni che potrebbero determinare alterazioni non compatibili con la vita. Fra questi gioca un ruolo fondamentale il glucosio, sostenitore della gran parte del metabolismo energetico delle cellule. Esso è controllato da numerosi ormoni tra cui quelli prodotti dal pancreas assumono una funzione predominante. Il pancreas, infatti, oltre al tessuto ghiandolare acinoso deputato alla produzione di succo pancreatico, comprende anche un tessuto endocrino preposto alla produzione di ormoni e costituito da piccole formazioni isolate denominate "Isole di Langerhans", dal nome dello scopritore. Qui trovano collocazione due tipi fondamentali di cellule con attività endocrina, disposte in ciascuna isola secondo un ordinamento preciso: le cellule β o B, addensate nella porzione centrale, le cellule α o A, dislocate alla periferia dove costituiscono una specie di zona corticale. Entrambe sono ricche di granuli contenenti quegli ormoni, elaborati in forma di precursori, che regolano la glicemia, ovvero il livello ematico di glucosio. Nello specifico, le cellule α secernono glucagone, mentre le cellule β producono insulina. Il primo, definito l'ormone "della carestia", è composto da un'unica catena peptidica lineare

di 29 amminoacidi ed ha un effetto iperglicemizzante: infatti il glucagone ha lo scopo di innalzare la glicemia all'interno di un organismo, esaltando i due processi fondamentali che portano all'immissione di glucosio all'interno del sangue, ovvero la glicogenolisi (durante la quale vengono degradate molecole di glicogeno fino ad ottenere il monosaccaride glucosio) e la neoglucogenesi (la quale determina la trasformazione di un composto non glucidico in glucosio). Il fattore regolatore principale del glucagone è rappresentato dalle variazioni di glicemia: quando vi è una diminuzione del livello ematico di glucosio (ad esempio in occasione di un digiuno o di un lavoro muscolare intenso), le cellule α sono potentemente stimolate nell'aumentare la produzione dell'ormone; viceversa, quando vi è un aumento (ad esempio in occasione di un pasto), la sua produzione viene inibita. In quest'ultimo caso, infatti, in opposizione al lavoro svolto dal glucagone, è presente nell'organismo l'insulina, detta anche ormone "dell'abbondanza". Essa è composta da due catene polipeptidiche lineari per un totale di 51 amminoacidi e si considera essere uno degli ormoni fondamentali nel mantenimento dell'omeostasi metabolica, ma la manifestazione più evidente della sua azione è sicuramente nel metabolismo glucidico. Qui ha un effetto ipoglicemizzante: si occupa di favorire l'abbassamento della glicemia, esaltando i processi che portano ad una sottrazione di glucosio dal sangue (ad esempio aumentando l'assunzione del glucosio stesso da parte delle cellule) ed inibendo quelli che portano ad una sua immissione. Agevola, quindi, meccanismi come la trasformazione del glucosio in glicogeno o in grassi oppure la demolizione ossidativa della molecola glucidica, accrescendo da un lato le riserve di materiale energetico dell'organismo e dall'altro la liberazione metabolica di energia. La regolazione dell'insulina è gestita dalle stesse cellule β che la producono, le quali fungono da sensori per la rilevazione del livello ematico di glucosio: la diminuzione della glicemia inibisce la liberazione dell'ormone, l'aumento la stimola. In particolare, la definizione del preciso livello di insulinemia viene fornita dal dosaggio del peptide C: è un peptide di collegamento che unisce l'amminoacido terminale della catena A e quello terminale della catena B all'interno del precursore dell'ormone prodotto dalle cellule β del pancreas, la proinsulina. Quando quest'ultima si scinde, contemporaneamente alla liberazione di insulina, si ha il distacco del peptide C, il quale viene rilasciato in quantità equimolari all'ormone. Proprio per questa sua caratteristica, nonostante non sia dotato di funzione biologica, è considerato la misura più fedele della quantità di insulina secreta dal pancreas, sia in condizioni basali che dopo stimoli fisiologici o farmacologici. Una lesione selettiva ed irreversibile delle cellule β porta alla soppressione della produzione di insulina da parte

dell'organismo e, di conseguenza, ad alterazioni significative del metabolismo glucidico, determinando una forma morbosa, frequente nell'uomo, nota come "diabete" [2-4-5].

1.2 FISIOPATOLOGIA

Il diabete mellito si manifesta nell'uomo con vari sintomi evidenti, quali l'aumento della diuresi (poliuria), l'aumento della sete (polidipsia) e della fame (polifagia), con una conseguente diminuzione del peso corporeo. Fra questi, il sintomo più precoce e caratteristico della patologia, provocato da carenza di insulina, è un persistente innalzamento della glicemia, la quale si presenta già elevata in condizioni di digiuno e cresce enormemente dopo i pasti. Di conseguenza, il glucosio, in difetto dell'ormone, non riesce a penetrare normalmente nelle cellule di gran parte dei tessuti, nonostante la sua concentrazione nel sangue sia elevata, provocando una sorta di "fame cellulare" (la mancata azione dell'insulina sui recettori, bersaglio della sua azione, impedisce alle cellule di ricevere il segnale dell'avvenuto pasto e, quindi, della presenza di sostanze energeticamente utilizzabili, cosicchè esse sono poste nella stessa condizione in cui sarebbero se il cibo non fosse stato introdotto). Si ha quindi la paradossale condizione di un eccesso di glucosio extracellulare associato ad una carenza di glucosio intracellulare. In questo modo, il metabolismo, in carenza di disponibilità glucidica (di cui ha bisogno essendo la maggior fonte energetica dalla quale normalmente viene alimentato), si orienta verso l'utilizzo dei lipidi, che vengono prelevati dal tessuto adiposo in elevata quantità e, giunti ai vari tessuti, penetrano nelle cellule dove ha inizio la loro demolizione ossidativa. La problematica principale insorge durante quest'ultimo processo: la quantità di acetil coenzima A (anche detto "Acetil-CoA") prodotta dall'ossidazione degli acidi grassi non può essere completamente smaltita e l'eccesso subisce, soprattutto nel fegato, la trasformazione in corpi chetonici, acidi deboli la cui concentrazione nel sangue (e di conseguenza anche nell'urina) aumenta considerevolmente (con lo scopo di fornire energia alle cellule "affamate") fino a giungere ad uno stato di acidosi metabolica (i corpi chetonici spostano il pH plasmatico in senso acido). Il persistere di questo disordine può successivamente sfociare, in base alla gravità dei sintomi, in acidosi vera fino ad arrivare al coma chetoacidosico, la componente più pericolosa del coma diabetico (manifestazione che può essere mortale nell'uomo). L'innalzamento del tasso glicemico ha importanti riflessi anche sulla funzione renale. Infatti, se la concentrazione di glucosio nel sangue supera la soglia di escrezione renale, la porzione di esso che non può essere riassorbita dai tubuli

renali resta nel liquido intratubulare e viene escreta con l'urina, provocando, quindi, glicosuria. La presenza di glucosio nel liquido intratubulare, inoltre, si oppone al riassorbimento dell'acqua, causando la poliuria e contribuendo al ridotto riassorbimento di sali inorganici. In questo modo i tessuti, già carenti di apporto glucidico, sono anche disidratati e scarsamente irrorati dal sangue. Il disordine metabolico, infine, va anche ad esaltare il processo di demolizione delle proteine, causando un dimagrimento notevole dell'individuo per via dell'impoverimento strutturale delle cellule. Nel soggetto diabetico, quindi, l'intera economia organica è perturbata nella sua globalità con conseguenze varie, quali la caduta delle difese immunitarie oppure alterazioni del sistema nervoso (come le neuropatie cardiache che colpiscono più frequentemente i nervi periferici) e di particolari formazioni come la retina dell'occhio (retinite diabetica) [2-4].

1.3 DIAGNOSI

Il diabete mellito, come abbiamo visto, rappresenta quindi una patologia estremamente eterogenea basata sulla perdita progressiva della secrezione insulinica, conseguenza dell'insulino-resistenza, o sulla completa autodistruzione delle cellule β , con successiva carenza assoluta dell'ormone nell'organismo. Nel continuum fra la patologia diabetica (DM) e la normale tolleranza al glucosio (NGT), però, si inserisce un'ulteriore categoria di pazienti che, nonostante abbia valori metabolici normali, presenta già un'iniziale deficienza della funzione insulinica. Questa condizione, nota come "prediabete", non è considerata un'entità clinica a sè ma piuttosto come un possibile fattore di rischio di evoluzione verso il diabete conclamato. In particolare, nel 1997 e nel 2003 un comitato di esperti nella diagnosi e nella classificazione del diabete mellito, all'interno di questa categoria, ha identificato due situazioni metaboliche differenti, caratterizzate da valori di glicemia superiori alla norma, ma inferiori a quelli che definiscono il diabete: ridotta tolleranza al glucosio (Impaired Glucose Tolerance, IGT), con livelli anormali di tolleranza al glucosio, e IFG (Impaired Fasting Glucose), con alterati livelli di glicemia a digiuno. In entrambi i casi è dimostrabile la presenza di un deficit nella secrezione insulinica e di una riduzione alla sensibilità all'ormone con, tuttavia, modalità ed intensità differenti. I soggetti IGT (per la maggior parte donne in età avanzata) dimostrano un'elevata resistenza all'ormone a livello muscolare e presentano una evidente disfunzione β -cellulare, sia nella fase precoce della risposta insulinica sia in quella tardiva: infatti, rispetto ai soggetti NGT, mantengono elevata la glicemia ancora 2 ore dopo l'assunzione di glucosio, avvenuta per

via orale o attraverso un pasto. La categoria degli IFG (in gran parte uomini in età media), invece, riscontra un'alta resistenza insulinica a livello epatico che contribuisce in maniera determinante alla caratteristica distintiva presente in questi soggetti, ovvero l'aumento della glicemia a digiuno. Mantiene un livello glicemico elevato principalmente entro i 30-60 minuti dall'assunzione di glucosio (con valori molto più elevati dei soggetti IGT), ma con un successivo ritorno rapido alla glicemia normale entro 2 ore. La diminuzione della fase precoce della risposta insulinica ha la sua causa nella cecità delle cellule β che, in questa situazione, non si manifesta solo come una ridotta sensibilità di queste ai valori di glicemia, ma anche come una diminuita sensibilità alle variazioni del glucosio stesso. In ogni caso, per svelare qualsiasi eventuale alterazione del metabolismo glucidico, è necessario sottoporre il soggetto ad indagini dinamico-funzionali, fondamentali nella distinzione fra soggetti normali (NGT), con ridotta tolleranza al glucosio (IGT o IFG) e diabetici (DM): dette indagini sfruttano la modificazione di uno o più parametri, indici della situazione metabolica del paziente, in condizioni di stimolo provocato e misurato, al fine di creare una condizione di stress metabolico riportabile a quella che l'individuo riscontra nella sua vita quotidiana. In particolare, i parametri più utilizzati sono la glicemia e l'insulinemia, considerate singolarmente o contrapposte. Un esempio fondamentale di queste prove dinamico-funzionali è l'utilizzo della curva da carico del glucosio nell'OGTT, ovvero "Orale Glucose Tolerance Test": si basa sul principio che lo studio della glicemia, dopo l'assunzione per via orale di una dose standard di glucosio, è indice del sistema insulare del pancreas e, più specificamente, della secrezione insulinica. Attraverso l'utilizzo della prova da carico, quindi, si va a vedere il livello di iperglicemia cui l'individuo va incontro in seguito a vari stimoli fisiologici: il soggetto diabetico e quello con ridotta tolleranza al glucosio, a differenza di quello con normale tolleranza, presentano un rischio di malattia proporzionale ai livelli iperglicemici che raggiungono, spia del difetto di funzionamento dell'insulina e quindi della gravità della malattia. Prima di eseguire una prova da carico di glucosio è opportuno, però, che il laboratorio accerti il valore della glicemia di base del soggetto: se questa è sufficientemente elevata, tanto da essere di per sé significativa per la diagnosi di diabete, diventa superfluo eseguire l'indagine. Bisogna inoltre controllare che il paziente abbia normali riserve di glicogeno, per cui, nei giorni precedenti la prova, lo stesso deve trovarsi in una condizione di dieta "normale", ovvero comprendente almeno 250 g di carboidrati al giorno (se così non fosse, bisogna somministrare per tre giorni una dieta normale con tale quantità di carboidrati). Stabilito ciò, la prova si esegue al mattino a digiuno, senza che il paziente abbia fumato o

bevuto, dopo averlo pesato e dopo averne misurato la temperatura (la prova non viene eseguita in presenza di febbre). Inoltre, il paziente deve essere riposato e deve rimanere seduto o sdraiato, senza mai alzarsi, per tutta la durata del test. Il glucosio viene sciolto in acqua, in una concentrazione di circa il 50%, eventualmente con aggiunta di succo di limone per facilitarne l'ingestione. La quantità da somministrare varia a seconda degli autori del test: viene, però, sempre più accettata come riferimento la quantità di 1 g/Kg che permette di ottenere una dose la cui totalità varia dalla metà ad un terzo della quota di carboidrati ingeriti in una dieta normale, durante un pasto. Il test avviene secondo un preciso ordine: un primo campione viene prelevato dopo un digiuno di almeno 12 ore; successivamente si fa ingerire il glucosio in un tempo non inferiore a 5 minuti; quindi si raccolgono i campioni di sangue ogni 30 minuti fino a 3, 4 o 5 ore, a seconda delle particolarità che si vogliono studiare. In conclusione, con i risultati ottenuti, si definisce la classe di appartenenza di ogni soggetto, in linea con le soglie diagnostiche stabilite dall'OMS:

- I soggetti DM sono definiti principalmente in base ai livelli di glicemia a digiuno, che deve essere ≥ 126 mg/dL, e/o su quelli di glicemia a 2 ore, che deve essere ≥ 200 mg/dL (in base al test diagnostico effettuato);
- La categoria degli IGT si presenta con una glicemia a 2 ore compresa fra 140 e 199 mg/dL;
- Gli IFG mostrano una glicemia a digiuno compresa fra 100 e 125 mg/dL.

Un'ulteriore metodologia, standardizzata in tempi più recenti, per effettuare la discriminazione fra le categorie sopra citate, è la misurazione dell'emoglobina glicata (HbA1c), ottenuta come prodotto della condensazione tra il glucosio e la catena β dell'emoglobina. Il comitato di esperti dell'American Diabetes Association ha individuato in un valore di HbA1c $\geq 6,5\%$ il cut-point diagnostico per il diabete mellito ed in un valore compreso fra 6 e 6,5% l'intervallo diagnostico per definire un soggetto con probabile alto rischio di diabete (realtà clinica non certa). Il principale vantaggio pratico di questa metodologia è rappresentato dalla non necessità, all'atto del prelievo di sangue, che il soggetto sia a digiuno (mentre per la glicemia è richiesto un digiuno di almeno 8 ore come già visto sopra). Allo stesso tempo, però, per definire un soggetto come affetto da patologia diabetica tramite dosaggio di HbA1c è necessario ripetere il test 2 volte, assicurandosi che le due prove siano effettuate nelle medesime condizioni [2-6-7-8].

1.4 CLASSIFICAZIONE

I soggetti con diabete conclamato sono tutti caratterizzati da valori insulemici ridotti; è possibile però fare una differenziazione fra coloro che hanno valori di insulina talmente bassi da richiedere una terapia sostitutiva per la regolazione della glicemia e coloro che invece possono utilizzare metodi alternativi (come ad esempio la dieta). Fra le due categorie di soggetti non esiste una demarcazione netta: non è possibile individuare un valore di insulinemia che possa funzionare da riferimento. Infatti, è difficile stabilire a che punto si possa parlare di insulino-dipendenza, rimandando la diagnosi a discrezione del medico che stabilirà la condizione del soggetto basandosi su un insieme di elementi clinico-laboratoristici. Occorre quindi effettuare una distinzione fra soggetti insulino-dipendenti e non insulino-dipendenti ai fini di una corretta terapia, nonostante la comparsa di complicanze incida ugualmente sui due gruppi. A queste si aggiungono altre forme di diabete, scoperte in tempi recenti ma diffusi in maniera già rilevante all'interno della popolazione [1-2].

1.4.1 Diabete di tipo 1

Il diabete mellito dipendente da insulina, anche definito “diabete di tipo 1”, colpisce più comunemente i soggetti nell'età della pubertà ed è caratterizzato da una carenza assoluta di insulina per via di una necrosi massiva delle cellule β . Questa può essere attribuita ad un processo autoimmune diretto contro le cellule β , verificatosi in individui in gran parte geneticamente predisposti, probabilmente a causa di un innesco ambientale. In seguito alla distruzione di queste cellule, il pancreas non risponde all'ingestione di glucosio e nell'organismo iniziano ad essere evidenti i sintomi dovuti alla mancanza di insulina. In un individuo normale, dopo l'ingestione di un pasto, si ha, entro due minuti e per un totale di quindici minuti, una scarica di secrezione di insulina in risposta al transitorio aumento dei livelli circolanti di glucosio nel sangue. Il soggetto con diabete, invece, non avendo cellule β funzionanti, non può gestire la glicemia a livello ematico, né mantenere una secrezione anche basale di insulina. Per ovviare a questo problema, si ricorre all'assunzione di insulina somministrata tramite microinfusori o penne apposite per controllare l'iperglicemia e per evitare la chetoacidosi, pericolose per la vita del paziente. Si mantengono, inoltre, costanti le concentrazioni ematiche di glucosio attraverso l'aiuto di misuratori portatili, facilitando il controllo diretto da parte del paziente che, in questo caso,

esegua da sé il trattamento mediante iniezione di ormone, oppure della cosiddetta “pompa di insulina”, ovvero l’infusione sottocutanea continua dell’ormone che rilascia il tasso basale di secrezione di insulina, permettendo di ovviare alle somministrazioni multiple quotidiane [1].

1.4.2 Diabete di tipo 2

La maggior parte dei soggetti con diabete appartiene alla categoria dei soggetti affetti da diabete mellito non dipendente da insulina, anche definita “diabete di tipo 2”. La malattia è in questo caso influenzata da fattori genetici, dall’età, dall’obesità piuttosto che da reazioni autoimmuni. Nello specifico, il pancreas mantiene una certa attività delle cellule β , ma la secrezione variabile di insulina non è sufficiente per mantenere l’omeostasi del glucosio. Inoltre la patologia è accompagnata frequentemente da una mancata sensibilità degli organi bersaglio all’insulina, considerata la causa principale di questa tipologia di diabete. Per mantenere le concentrazioni ematiche di glucosio entro limiti normali, così da ridurre l’insulino-resistenza, si possono effettuare modificazioni della dieta quotidiana o esercizio fisico. La maggior parte dei soggetti con diabete di tipo 2, però, dipende dall’intervento farmacologico di ipoglicemizzanti assunti oralmente oppure da una terapia insulinica, soprattutto durante la progressione della malattia, la quale porta ad una graduale ma continua riduzione delle cellule β [1].

1.4.3 MODY

Il MODY (“Maturity-Onset Diabetes of the Young”), noto anche come “diabete di tipo 3”, rappresenta circa l’1-5% di tutti i casi di diabete. Si tratta di un disordine ereditario autosomico che non produce corpi chetonici e si manifesta prima dei 25 anni. Le persone affette da questa forma di diabete non sono obese, non si riscontra in loro una resistenza all’insulina [1].

1.4.4 LADA

Noto anche come “diabete autoimmune latente negli adulti”, è una forma comune ma poco studiata della patologia diabetica, con caratteristiche ibride fra il diabete di tipo 1 e quello di tipo 2. Non esiste un’accordo uniforme sulla sua definizione, ma la diagnosi si basa solitamente su tre criteri:

- Età adulta all'esordio;
- Autoanticorpi delle isole come marker di attività autoimmune (caratteristica tipica del diabete di tipo 1 ma con un'evoluzione più lieve e con una progressione del fallimento delle cellule β più lenta);
- Indipendenza dall'insulina (particolarità evidenziata nel diabete di tipo 2 ma con una minore secrezione insulinica e con una progressione più rapida verso la dipendenza dall'ormone) [9].

1.4.5 Diabete gestazionale

Classificato come “diabete di tipo 4”, è definito come un'intolleranza al glucosio associata alla gravidanza. Il feto, infatti, utilizza in maniera preferenziale il glucosio materno, diminuendo lo stimolo della secrezione di insulina nell'organismo della madre. Questa condizione, legata alla situazione di momentaneo stress, comporta la necessità di un trattamento dietetico-farmacologico come per il diabete primario ma rappresenta uno stato di salute reversibile, tanto che questi soggetti vengono poi rivalutati alla fine della gravidanza. Le donne a cui è stata diagnosticata questo tipo di patologia diabetica, però, sono ad elevato rischio di sviluppare il diabete di tipo 2 entro 5 anni dal parto: il diabete gestazionale, infatti, è considerato come uno stadio precoce nella storia del diabete mellito di tipo 2, in quanto presentano le stesse caratteristiche, come l'aumento dell'insulino-resistenza e la riduzione della produzione insulinica [1-2].

2. ANALISI CON TECNICHE DI CLUSTERING DI CURVE DA TEST DI TOLLERANZA AL GLUCOSIO PER VIA ORALE

2.1 Stato dell'arte

Lo sviluppo del diabete di tipo 2 è un processo lento e la sua manifestazione è spesso preceduta, come abbiamo già detto in precedenza, da una fase di prediabete, frequentemente non diagnosticata. Tuttavia, solo circa il 50% dei soggetti con regolazione glicemica compromessa (quindi in fase di prediabete) progredisce verso la patologia conclamata. Pertanto, la determinazione dei sottotipi del prediabete e del diabete diventa essenziale per la prevenzione personalizzata della patologia. Per ovviare a questo problema, negli ultimi tempi, sono state impiegate tecniche di “Machine Learning”, le quali hanno lo scopo di costruire modelli statistici da set di dati estremamente vasti. Fra queste, maggiormente utilizzata nella discriminazione dei soggetti fra NGT, IGT/IFG e DM è la tecnica del “clustering”: partendo da variabili derivate dalle curve glicemiche ottenute durante il test di tolleranza al glucosio, i soggetti vengono raggruppati in modo tale che dati appartenenti allo stesso cluster condividano caratteristiche simili, viceversa dati appartenenti a cluster diversi siano differenti fra loro. In particolare, tra le tecniche di clustering più utilizzate in questo ambito, vi sono il “k-means”, che divide i dati in k gruppi basando la partizione sulla creazione di un centroide per ogni cluster e restituendo una funzione che assegna ogni dato ad uno dei cluster, ed il “clustering gerarchico”, che mira alla formazione di una gerarchia fra i cluster, partizionando i dati con una procedura di ottimizzazione sequenziale [10-11-12].

2.1.1 Clustering basato su OGTT proposto da Wagner et al.

Il seguente studio ha come scopo quello di dimostrare che la stratificazione delle popolazioni ad aumentato rischio di diabete di tipo 2, utilizzando semplici caratteristiche cliniche (variabili derivate dal test di tolleranza al glucosio orale), consente strategie precise ed efficienti nella prevenzione della patologia stessa. Il clustering, pur non fornendo sottotipi definitivi per ogni singolo paziente, può essere infatti utile per

caratterizzare l'eterogeneità metabolica prima della manifestazione clinica del diabete di tipo 2 e per suggerire potenziali implicazioni terapeutiche [11].

Lo studio mirava ad identificare i sottotipi che riflettono le differenze nei processi fisiopatologici della storia naturale del diabete di tipo 2. Il clustering iniziale è stato effettuato utilizzando i dati di un sottogruppo di partecipanti (n=899) dello studio TUEF/TULIP che non avevano valori mancanti per le variabili di fenotipizzazione preselezionate: glicemia; sensibilità all'insulina e sua secrezione definite attraverso indici basati sull'OGTT (caratteristiche chiave in quanto il paradigma principale della patogenesi della malattia è un insufficiente aumento compensatorio della secrezione in risposta all'insulino-resistenza); colesterolo HDL (noto da tempo come elemento esplicativo della sindrome metabolica e dell'insulino-resistenza); volume di grasso epatico, sottocutaneo e viscerale; punteggio di rischio poligenico per il diabete di tipo 2, per valutare la responsabilità genetica (così da differenziare meglio la disfunzione delle cellule β geneticamente determinata da quella provocata dall'ambiente). Per ottenere queste caratteristiche, a tutti i partecipanti è stata somministrata una soluzione di glucosio da 75g alle ore 8:00 (dopo un digiuno notturno) ed è stato inserito un catetere (prima e 30,60,90 e 120 minuti dopo l'ingestione di detta soluzione) al fine di ottenere un campione di sangue venoso. Il clustering è stato successivamente replicato nella coorte Whitehall II (n=6810) utilizzando variabili concettualmente simili (variabili cosiddette "proxy", in quanto caratteristiche come il grasso epatico, viscerale e sottocutaneo non erano disponibili in questa coorte): glicemia durante il test del glucosio; sensibilità dell'insulina e sua secrezione; insulina e trigliceridi a digiuno; circonferenza della vita e dell'anca; indice di massa corporea (BMI); colesterolo HDL. Affinchè la riassegnazione dei cluster avvenisse in maniera efficiente e precisa, sono stati calcolati i valori mediani delle variabili proxy nelle due coorti e, in Whitehall, i cluster sono stati scelti in base alle distanze euclidee dai valori ottenuti in TUEF/TULIP. In conclusione, sono stati identificati 6 cluster impiegando 8 variabili nella coorte TUEF/TULIP e gli stessi sono stati replicati nella coorte Whitehall impiegando 9 variabili (Tabella 1) [11].

(Tabella 1:Descrizione delle caratteristiche principali dei 6 cluster definiti in entrambe le coorti)

Cluster	Caratteristiche principali	Obesità e Distribuzione del Grasso	Sensibilità all'insulina	Secrezione di insulina	Glicemia	Altre caratteristiche specifiche
1	Rischio basso	Sovrappeso	Media	Adeguate	Prevalentemente NGT	/
2	Rischio molto basso	Normale	Buona	Adeguate	Prevalentemente NGT	/
3	Fallimento delle cellule β	Sovrappeso/Obeso	Moderatamente Bassa	Bassa	Prevalentemente Prediabete	Aumentato rischio genetico di diabete di T2DM
4	Obesità a basso rischio	Obeso	Buona	Adeguate	Prevalentemente NGT	/
5	Insulino-resistenza ad alto rischio	Obeso	Molto Bassa	Bassa	Prevalentemente Prediabete (la maggior parte IGT con o senza IFG)	Rischio genetico di T2DM sopra la media e grasso epatico molto alto
6	Nefropatia adiposa viscerale ad alto rischio	Obeso	Moderatamente Bassa	Moderatamente Bassa	NGT e Prediabete (la maggior parte IFG)	Rischio genetico di T2DM basso, elevato grasso viscerale e del seno renale

Per il rilevamento della patologia diabetica è stato utilizzato uno dei seguenti metodi: diabete clinicamente accertato (dall'anamnesi del paziente o mediante l'uso di un farmaco per il diabete), elevato glucosio a digiuno (≥ 7 mmol/l), glucosio post-challenge $\geq 11,1$ mmol/l. In entrambe le coorti, l'incidenza della patologia è stata elevata nel cluster 5, subito seguito dal cluster 3. In TUEF/TULIP, i partecipanti del cluster 6 non hanno dimostrato un aumentato rischio di patologia, mentre in Whitehall era moderatamente

elevato. Al contrario, il cluster 2 in Whitehall presentava un rischio significativamente inferiore di sviluppare il diabete rispetto al cluster 1. In particolare, nella coorte Whitehall, relativamente pochi partecipanti hanno sviluppato la patologia nei cluster metabolicamente sani (cluster 1, 48 su 817; cluster 2, 62 su 2552; cluster 4, 14 su 314) e, fra questi, la maggior parte ha manifestato forme di diabete lieve correlato all'obesità e all'età. Nel cluster 6, invece, 13 dei 23 soggetti hanno sviluppato una forma grave di diabete insulino-resistente. Nello specifico:

- Il cluster 5 è stato identificato come la sottopopolazione con il più alto rischio di diabete di tipo 2: i soggetti di questo gruppo presentavano obesità, insulino-resistenza, alti livelli di grasso epatico e bassa secrezione di insulina;
- Il cluster 6 rappresentava un fenotipo insulino-resistente, in cui i partecipanti mostravano elevate quantità di grasso viscerale ma minore grasso epatico e maggiore secrezione di insulina rispetto al cluster 5: circa la metà dei soggetti presentava prediabete al momento dell'arruolamento, tuttavia con glicemia media inferiore e rischio di patologia moderato rispetto al cluster 5 nonché con un rischio genetico relativamente basso;
- Il cluster 4 comprendeva individui con obesità ma con basso deterioramento glicemico, compatibile con il concetto di obesità metabolicamente sana: era associato ad un minor rischio di diabete di tipo 2, in quanto i soggetti presentavano grasso corporeo immagazzinato prevalentemente nei depositi sottocutanei, piuttosto che viscerali (modello noto per essere metabolicamente più favorevole);
- Il cluster 3 presentava un fenotipo caratterizzato da elevato rischio genetico e bassa secrezione di insulina (cause che potrebbero spiegare l'elevata incidenza della patologia in questo gruppo); inoltre, nonostante il basso indice di predisposizione come quello del cluster 5, mostrava una maggiore sensibilità all'insulina [11].

2.1.2 Clustering basato su OGTT proposto da Obura et al.

Utilizzando i dati di un OGTT campionato frequentemente (OGTT a sette punti) si è dimostrato che è possibile identificare i modelli della curva del glucosio associati a diversi profili clinici al basale e diversi tassi di successivo deterioramento clinico, così da poter dividere in sottogruppi le persone a rischio elevato di diabete di tipo 2 [12].

La ricerca ha incluso i dati di 2126 partecipanti al basale e 1988 a 18 mesi, dopo aver escluso quelli persi al follow-up, presi dallo studio IMI-DIRECT (studio europeo che

comprendeva individui a rischio elevato di deterioramento glicemico, provenienti da diverse coorti europee). Su ognuno di essi è stato eseguito un OGTT da 75g (al basale e a 18 mesi) nonché un prelievo di sangue a 0,15,30,45,60,90,120 minuti dopo un digiuno notturno di 10 ore. Sono stati poi determinati il glucosio plasmatico (mmol/l), il Peptide C (mmol/l), l'insulina (pmol/l) e l'HbA1c (mmol/mol). Infine sono stati calcolati l'indice di Matsuda, l'insulino-resistenza, l'indice glicemico e, a partire dal test di tolleranza al glucosio, è stato utilizzato un modello matematico che descriveva l'associazione tra la concentrazione di glucosio e la secrezione di insulina, così da poter stimare il tasso di secrezione totale dell'ormone. Attraverso l'uso della media stimata delle curve glicemiche ottenute tramite OGTT, sono stati identificati 4 sottogruppi della curva del glucosio dagli OGTT al basale ed al follow-up, etichettati in ordine crescente di livelli di picco di glucosio di 1 ora. Nello specifico, al basale:

- Il sottogruppo 1 conteneva 577 partecipanti e presentava il picco glicemico più basso, ma verificatosi prima rispetto a tutti gli altri cluster;
- Il sottogruppo 2 aveva 1012 soggetti caratterizzati dal picco glicemico più alto ad 1 ora;
- Il sottogruppo 3 presentava 327 partecipanti ed il picco glicemico più alto a 2 ore;
- Il sottogruppo 4 raccoglieva 210 individui con il picco glicemico più elevato ad 1 ora, simile a quello del sottogruppo 2 [12].

Al basale, con l'eccezione degli indici di Matsuda e insulino-genico, gli individui nei sottogruppi con picco glicemico più elevato (cluster 2-4) mostravano valori del tratto glicemico più alti e quindi un profilo del tratto glicemico meno favorevole rispetto al sottogruppo 1. Questi ultimi presentavano il livello di glucosio a digiuno e a 2 ore molto simile ed impiegavano lo stesso tempo per raggiungere il picco (45 minuti). Tuttavia, i partecipanti del sottogruppo 4 avevano un profilo del tratto glicemico peggiore, con un valore del picco di glucosio significativamente più alto e con tasso di secrezione totale di insulina, indice di Matsuda e indice insulino-genico inferiori rispetto ai soggetti appartenenti al sottogruppo 2. Infine, il sottogruppo 3, caratterizzato da un valore a 2 ore più elevato, mostrava livello di glucosio a 1 ora e tasso di secrezione totale di insulina significativamente più alti rispetto al sottogruppo 2, che aveva un livello di glucosio a digiuno simile, ma con un livello inferiore a 2 ore. Dopo un follow-up mediano di 18 mesi, con le stesse metodologie sopra descritte, sono stati identificati 5 sottogruppi e 57 partecipanti che avevano sviluppato il diabete (probabilmente raccolti nel sottogruppo 5).

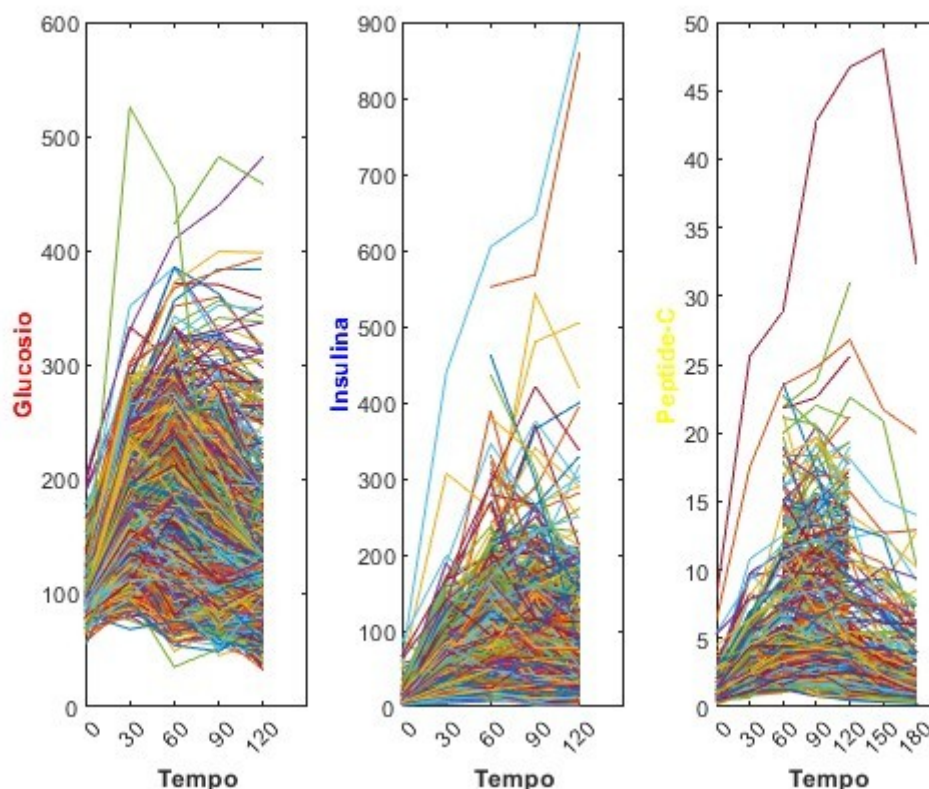
La variazione del tasso di secrezione totale di insulina, degli indici di Matsuda e insulinogenico si presentavano inferiori nei sottogruppi 2-4, indicando ridotta sensibilità all'insulina e ridotta secrezione della stessa rispetto al sottogruppo 1. Ugualmente è stato osservato confrontando i sottogruppi con lo stesso livello a digiuno ma diversi livelli di glucosio a 2 ore, ovvero 2 e 3, con i partecipanti di quest'ultimo che progredivano più velocemente. Infine, la maggior parte dei partecipanti è rimasta nello stesso sottogruppo o è passata a sottogruppi con picco più alto al follow-up. In conclusione, i partecipanti ai sottogruppi con picchi glicemici e livelli di glucosio a 1 ora più alti presentavano peggior profilo dei tratti glicemici rispetto ai soggetti del cluster con il picco più basso e più precoce. Allo stesso modo, il sottogruppo con il più alto picco di glucosio a 2 ore mostrava un profilo dei tratti glicemici peggiore rispetto ai partecipanti al cluster con livello di glucosio a digiuno simile, ma a 2 ore più basso. Inoltre, i sottogruppi con i peggiori profili glicemici presentavano anche i più alti livelli di glucosio a 1 ora, suggerendo che il glucosio plasmatico a 1 ora sia il predittore principale del rischio incidente di diabete di tipo 2 rispetto ad altre misure standard, come il glucosio a 2 ore. Sulla base degli attuali criteri dell'OMS per la diagnosi di alterata regolazione del glucosio, solo il 39% e il 63% dei partecipanti del sottogruppo 3 e il 51% e il 10% del sottogruppo 4, potevano considerarsi rispettivamente affetti da alterata glicemia a digiuno (IFG) e alterata tolleranza al glucosio (IGT) [12].

2.2 Definizione della popolazione di studio

Le analisi sono state condotte utilizzando un database disponibile presso il Dipartimento di Ingegneria dell'informazione e relativo a 2256 partecipanti: 1335 NGT, 741 IGT/IFG e 180 T2DM. Ognuno di loro è stato sottoposto ad OGTT dal quale sono stati ricavati il glucosio e l'insulina da 0 a 180 minuti (anche se nello studio sono valutati entrambi fino a 120 minuti per via della mancanza di dati) e il Peptide-C da 0 a 180 minuti.

2.3 Clustering

Per l'analisi con tecniche di clustering sono state utilizzate le curve di glucosio ed insulina da 0 a 120 minuti e quelli di Peptide-C da 0 a 180 minuti (per quest'ultimo, visto che molti pazienti presentavano dati mancanti, i partecipanti allo studio sono stati ridotti a 931) (Figura 1).



(Figura 1: Rappresentazione delle curve temporali del glucosio (da 0 a 120 minuti), dell'insulina (da 0 a 120 minuti) e del Peptide-C (da 0 a 180 minuti))

Attraverso la funzione “evalclusters”, supportata dalla funzione “rng” che garantisce la riproducibilità dei risultati, sulla base delle curve temporali, è stato definito un numero finito di cluster: è stato impiegato l’algoritmo “K-Means” secondo tre metodologie di clustering differenti (“Calinski”, “Davies”, “Silhouette”) per individuare il giusto numero di cluster secondo cui suddividere i dati a disposizione in base ai valori del glucosio (da 0 a 120 minuti), dell’insulina (da 0 a 120 minuti), dell’unione di questi ultimi due, del Peptide-C (da 0 a 180 minuti) ed infine in base ai valori di tutte e tre le caratteristiche presi insieme (naturalmente per quest’ultimo sono stati valutati esclusivamente i soggetti che possedevano tutte e tre le tipologie di valori). Il metodo Calinski-Harabasz, anche detto “del rapporto di varianza (VRC)”, si basa sul concetto che i cluster ben definiti presentano una grande varianza fra loro (SSB) ed una piccola varianza all’interno del cluster stesso (SSW): il rapporto fra quest’ultime mi definisce l’indice di Calinski-Harabasz (VRCK), la soluzione con il valore più alto di quest’ultimo definisce il numero ottimale di cluster per

la popolazione analizzata [13]. Il metodo Davies-Bouldin si basa sul rapporto tra le distanze all'interno del cluster e fra i cluster: la soluzione di cluster ottimale ha il valore dell'indice di Davies-Bouldin (DB) più basso [14]. Infine, il metodo Silhouette definisce il valore di silhouette (da -1 a 1) per ogni osservazione a disposizione, misurando quanto quel punto sia simile agli altri nello stesso cluster, rispetto ai punti negli altri cluster: un valore dell'indice di silhouette alto indica che l'osservazione è ben abbinata al proprio cluster e scarsamente agli altri, se la maggior parte dei punti di un cluster presenta elevato questo indice, allora la soluzione di clustering è appropriata [15]. Una volta determinato il numero esatto di sottogruppi, è stata utilizzata la funzione matlab "k-means" (sempre supportata dalla funzione "rng" per garantire la riproducibilità dei risultati) per affiancare ad ogni soggetto incluso nello studio il corrispettivo indice di cluster.

2.4 Variabili selezionate

Per il seguente studio sono state valutate variabili antropometriche sia qualitative (come il gruppo di appartenenza, indicante o meno la patologia) e quantitative (età, peso, altezza, indice di massa corporea BMI), introduttive per ogni soggetto; caratteristiche legate al rischio cardiovascolare (unito indissolubilmente alla patologia diabetica, in quanto questa determina un aumento da 2 a 4 volte superiore di sviluppare malattie cardiovascolari), come pressione sistolica e diastolica (misurate in mmHg), colesterolo totale, trigliceridi (TG) e lipoproteine sia ad alta che a bassa densità, HDL e LDL rispettivamente (calcolati in mg/dl); variabili glicemiche come l'HbA1C (misurata in % , permette di determinare la quantità di emoglobina glicata, ovvero la quota di emoglobina che si lega al glucosio), la glicemia da 0 a 120 minuti (mg/dl), ponendo particolare attenzione al valore di glucosio basale, l'AUCg e il deltaAUCg (misurati in mg/dl, indicano rispettivamente l'area sotto la curva del glucosio e l'area sotto la curva del glucosio soprabasale, durante un OGTT di 2 ore); caratteristiche tipiche della secrezione insulinica come l'insulinemia da 0 a 120 minuti ($\mu\text{U/ml}$), evidenziando nello specifico l'insulina basale, l'AUCi e il deltaAUCi (calcolati in $\mu\text{U/ml}$), rappresentano rispettivamente l'area sotto la curva dell'insulina e l'area sotto la curva dell'insulina soprabasale, durante un OGTT di 2 ore, l'indice insulinogenico, basato sull'area sottostante la curva totale, l'indice di deltaAUC, valutato sull'area sottostante la curva soprabasale totale, e l'indice insulinogenico classico IGI, dato dal rapporto fra la concentrazione di insulina a 30 minuti meno quella basale e la medesima differenza calcolata per il glucosio (misurati in pmol/mmol) ; features tipiche della sensibilità all'insulina, come l'indice "Oral Glucose Insulin Sensivity", noto come

OGIS, e l'indice OGIS modificato, l'indice "Quantitative Insuline Sensivity Check Index, definito QUICKI, (misurato in $w/\text{Log}10$ ed ottenuto utilizzando l'inverso della somma dei logaritmi dell'insulina e del glucosio basali) e l'indice di Matsuda; variabili specifiche della patologia diabetica come i "Disposition Index", ovvero DI1, DI2 e DI3, ottenuti moltiplicando l'OGIS modificato rispettivamente per l'AUC_i, il ΔAUC_i e per l'indice insulinogenico (incrociano la secrezione e la sensibilità insulinica per valutare la relazione iperbolica tra le due, andando ad analizzare la curva risultante); infine l'indice "Triglyceride Glucose", noto come TYG, caratteristico del rischio cardiometabolico, calcolato in questo studio con la formula: $(\ln(\text{Trigliceridi} * \text{Glucosio basale})) / 2$.

2.5 Analisi Statistica

Una volta ottenuta la clusterizzazione sulla base delle curve temporali, però, si presentava l'esigenza di capire che tipologia di soggetti e con quali caratteristiche fossero racchiusi in ogni raggruppamento. Una volta stabilito, attraverso un test di Lilliefors, che le variabili selezionate presentavano una distribuzione non normale, si è allora proseguito con il test di kruskalwallis, attraverso il quale tutte le features sono state suddivise in base ai cluster definiti in precedenza nei vari sottogruppi, così da determinare la tipologia di soggetti e relative caratteristiche contenuti in ogni singolo cluster. Lo stesso test ha fornito anche una struttura contenente le statistiche dei dati analizzati, fra le quali il P-value di ogni raggruppamento. Attraverso la funzione matlab "multcompare", è stato poi effettuato un test di confronto multiplo a coppie fra P-value di cluster diversi per ogni singola variabile selezionata, così da individuare quale fosse particolarmente significativa nella differenza fra i vari raggruppamenti e con la funzione "grpstats" sono state calcolate mediana e range interquartile (25°-75°) per tutte le variabili di ogni singolo cluster.

Infine, per rendere ancora più evidenti le differenze fondamentali tra i sottogruppi, i cluster sono stati rappresentati attraverso l'utilizzo di una standardizzazione (z-score) dei valori di ogni singola variabile prescelta: usufruendo della mediana e della deviazione assoluta mediana dell'intera popolazione, è stato definito il livello z-score per ogni singola osservazione di tutte le variabili scelte; successivamente per ogni variabile è stata calcolata la mediana dei livelli z definiti in precedenza. Ogni feature di ogni cluster è stata quindi identificata attraverso un livello z-standardizzato così da permetterne il confronto con le altre features a disposizione.

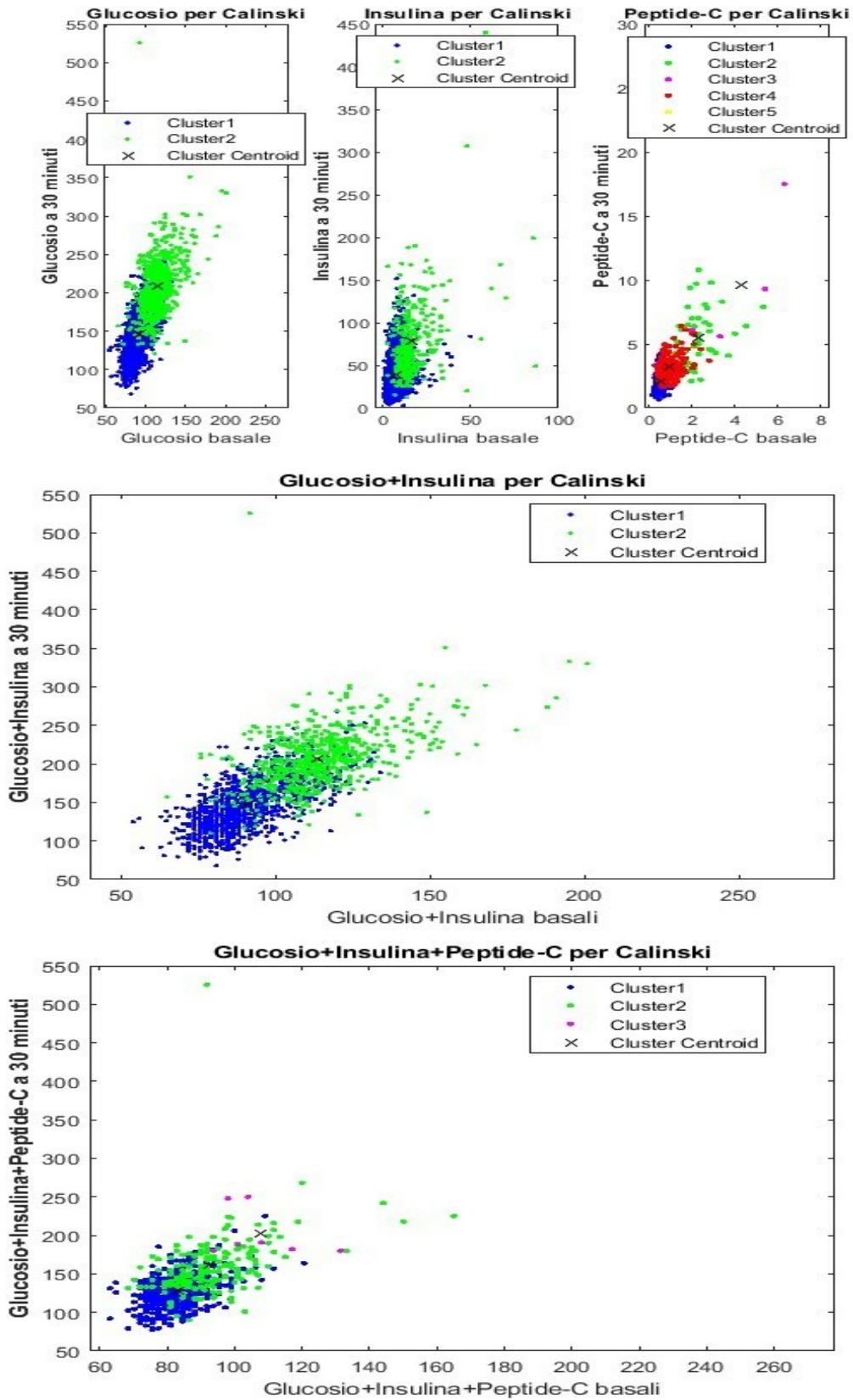
2.6 Risultati

Una volta definite le curve temporali del glucosio (da 0 a 120 minuti), dell'insulina (da 0 a 120 minuti) e del Peptide-C (da 0 a 180 minuti), si è proceduto con la custerizzazione delle stesse. Dalla tabella 2 si evince che in base alle curve glicemiche il numero di cluster ottimi è pari a 2 per tutte e tre le metodologie, mentre si riscontrano numeri di cluster differenti negli altri casi (Tabella 2).

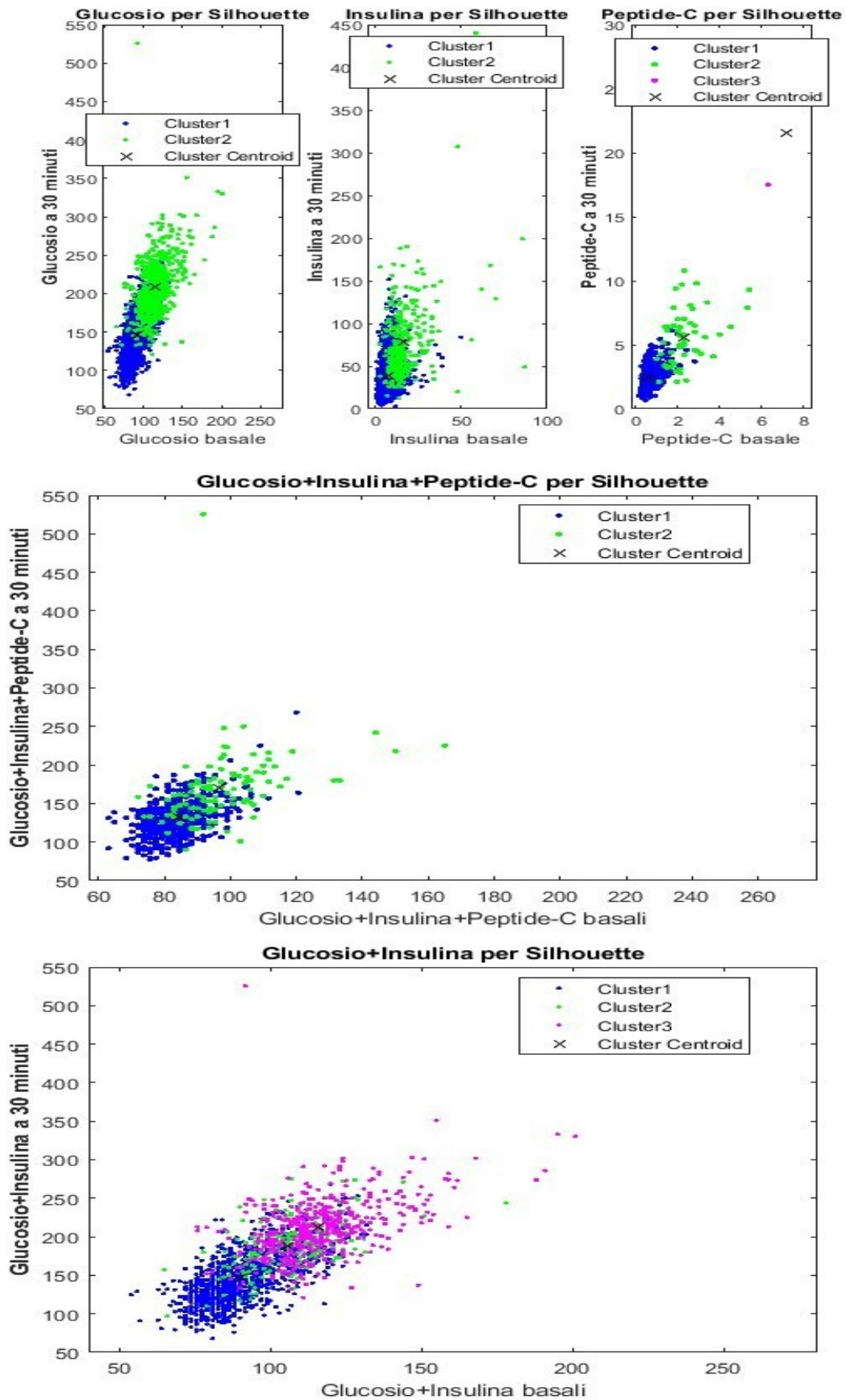
(Tabella 2: soluzione di clusterizzazione ottima in base ai 3 metodi selezionati (colonne) ed in base a glucosio, insulina, Peptide-C, glucosio ed insulina insieme e tutti e 3 i parametri scelti contemporaneamente)

TUTTE LE CATEGORIE			
	SILHUOETT E	CALINSKI	DAVIES
GLUCOSIO	2	2	2
INSULINA	2	2	3
PEPTIDE C	3	5	3
GLUCOSIO+INSULINA	3	2	2
TUTTI PARAMETRI	2	3	6

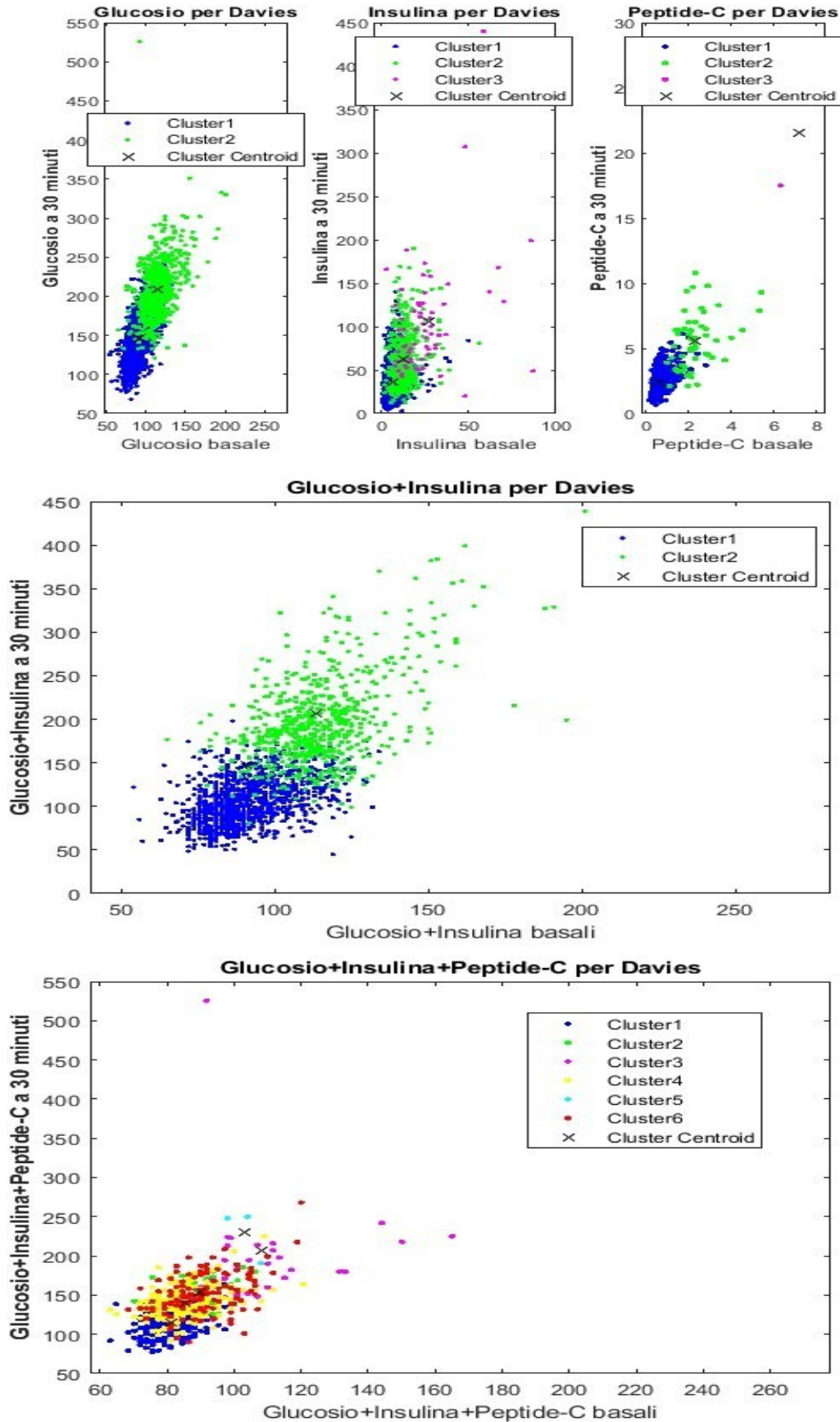
Successivamente, sono stati eseguiti i plot dell'intera popolazione suddivisa secondo la soluzione ottima di clustering per i 3 metodi ed in base alla variabile scelta (glucosio, insulina, entrambe, Peptide-C, tutte le variabili insieme). Nello specifico, sono stati confrontati per i cluster a disposizione specifici valori temporali di glucosio, insulina e Peptide-C: valore basale (al tempo zero) e valore dopo 30 minuti (Figura 2-3-4).



(Figura 2: sono rappresentati i plot della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Calinski in base ai valori della variabile scelta)

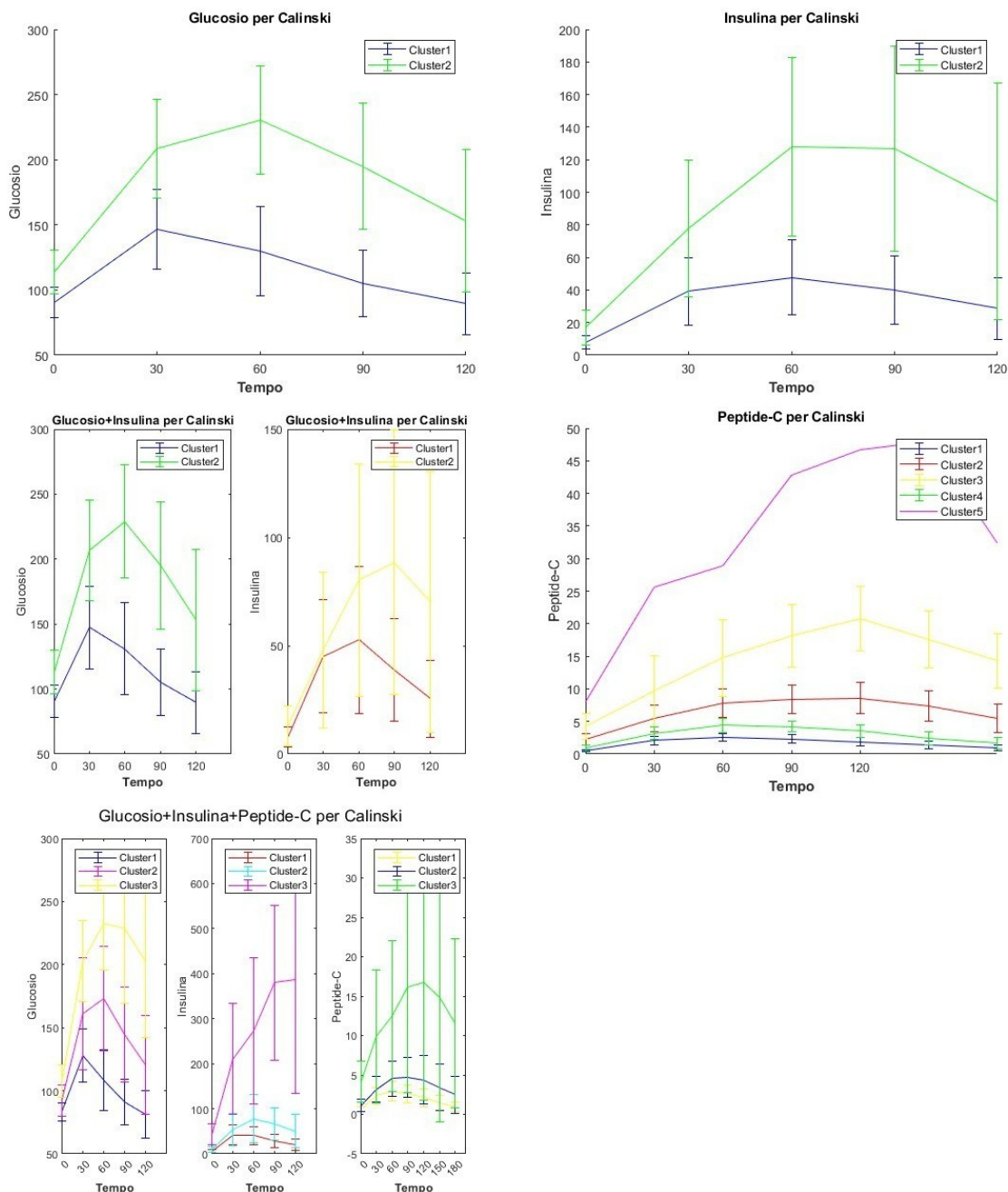


(Figura 3: sono rappresentati i plot della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Silhouette in base ai valori della variabile scelta)

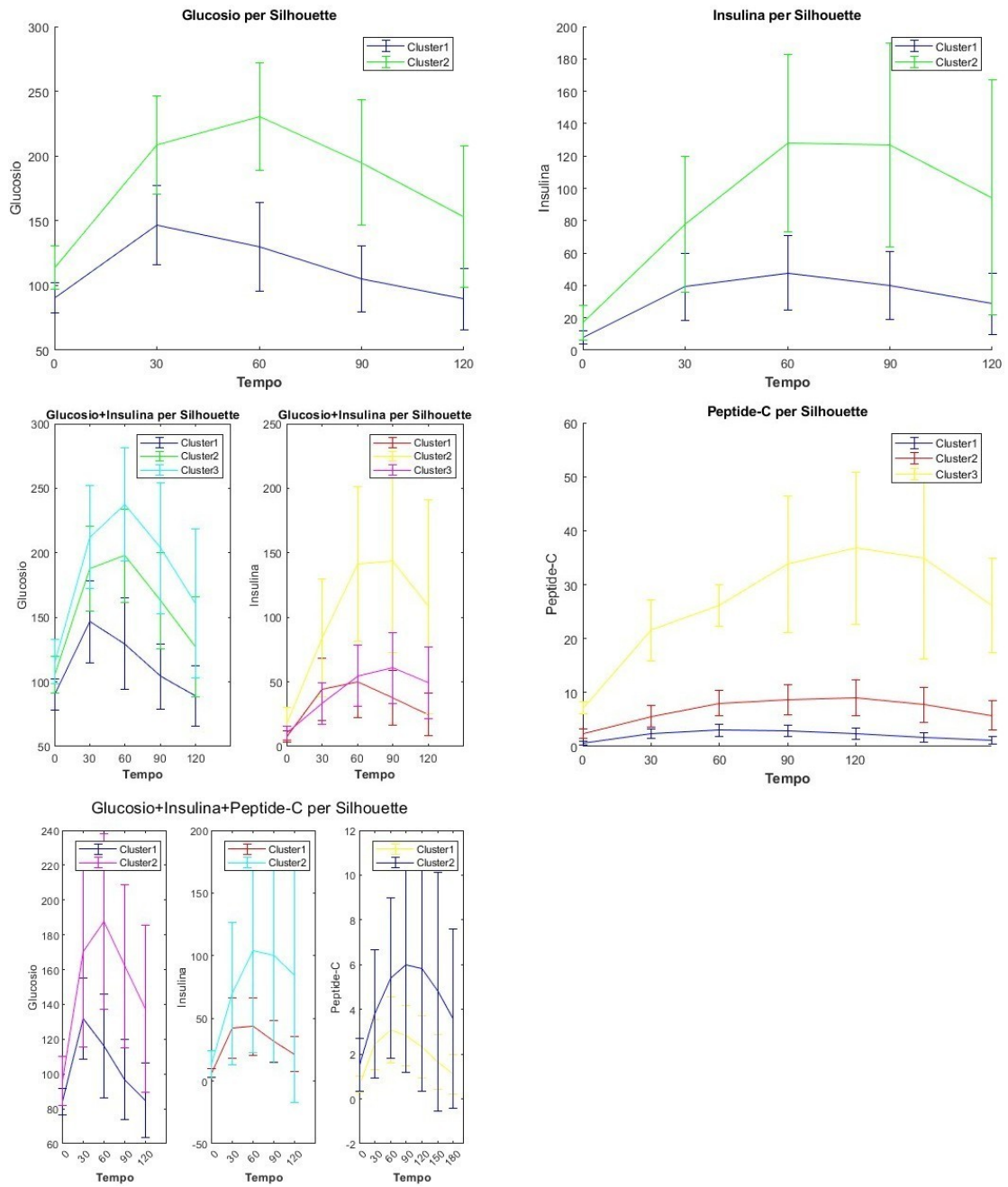


(Figura 4: sono rappresentati i plot della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Davies in base ai valori della variabile scelta)

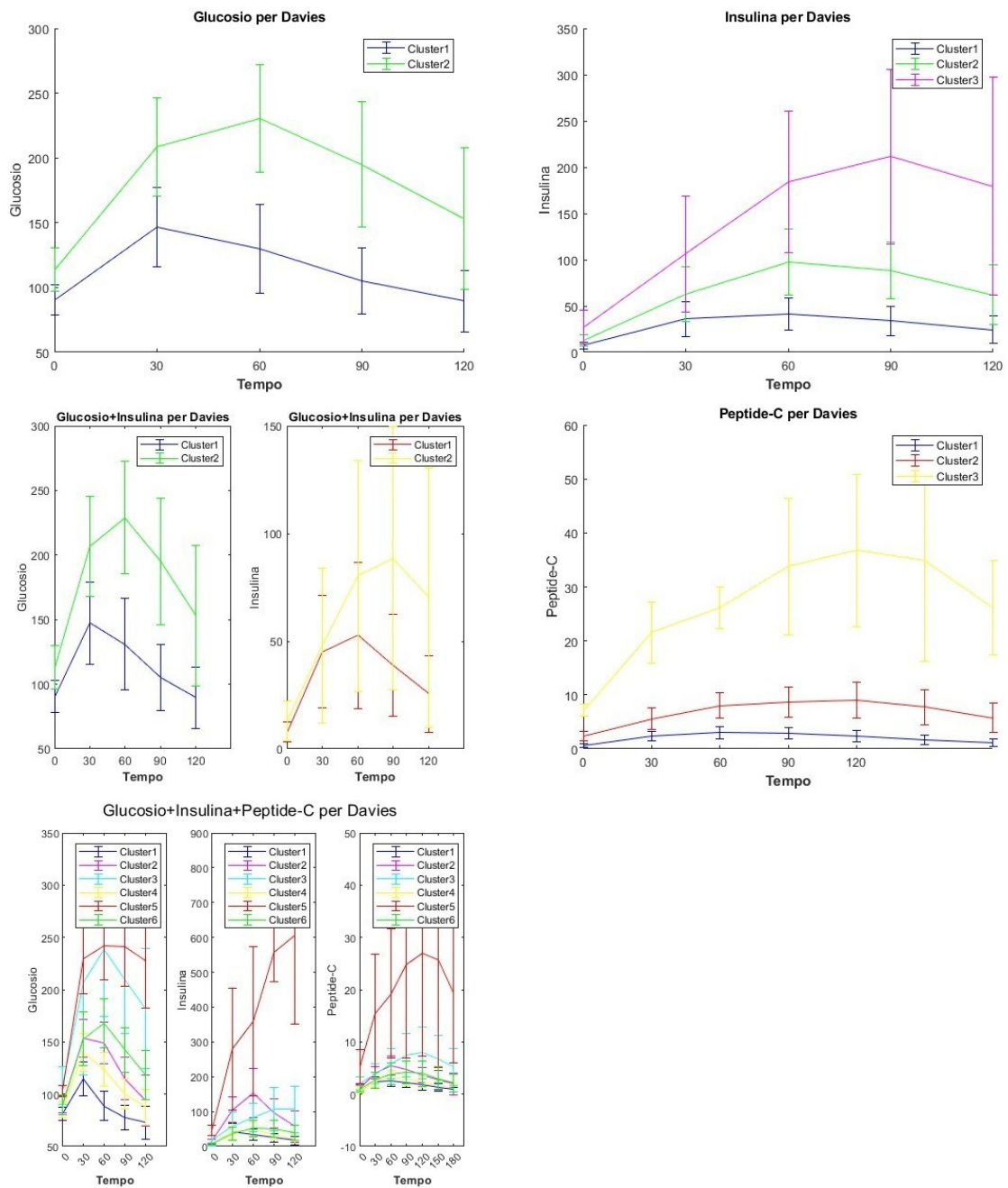
È stata utilizzata la funzione matlab “errorbar” per definire plot linearizzati dell’intera popolazione, per tutti e 3 i metodi e in base alla variabile scelta (glucosio, insulina, entrambe, Peptide-C, tutte le variabili), così da rendere evidente la differente shape delle curve temporali fra i cluster individuati (Figura 5-6-7).



(Figura 5: sono rappresentati i plot linearizzati della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Calinski in base ai valori temporali della variabile scelta)



(Figura 6: sono rappresentati i plot linearizzati della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Silhouette in base ai valori temporali della variabile scelta)

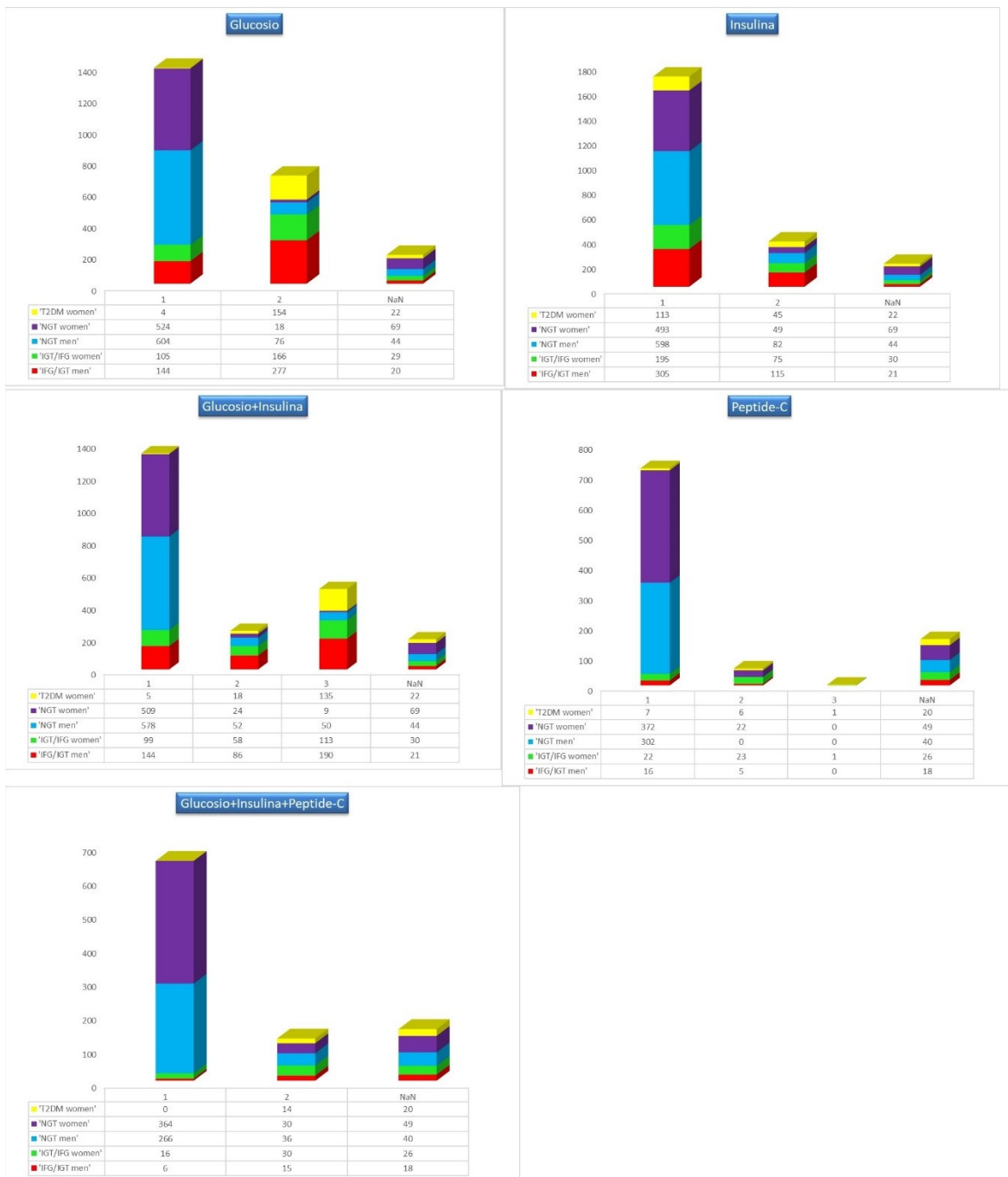


(Figura 7: sono rappresentati i plot linearizzati della popolazione divisa in base alla soluzione ottima di clustering secondo il metodo Davies in base ai valori temporali della variabile scelta)

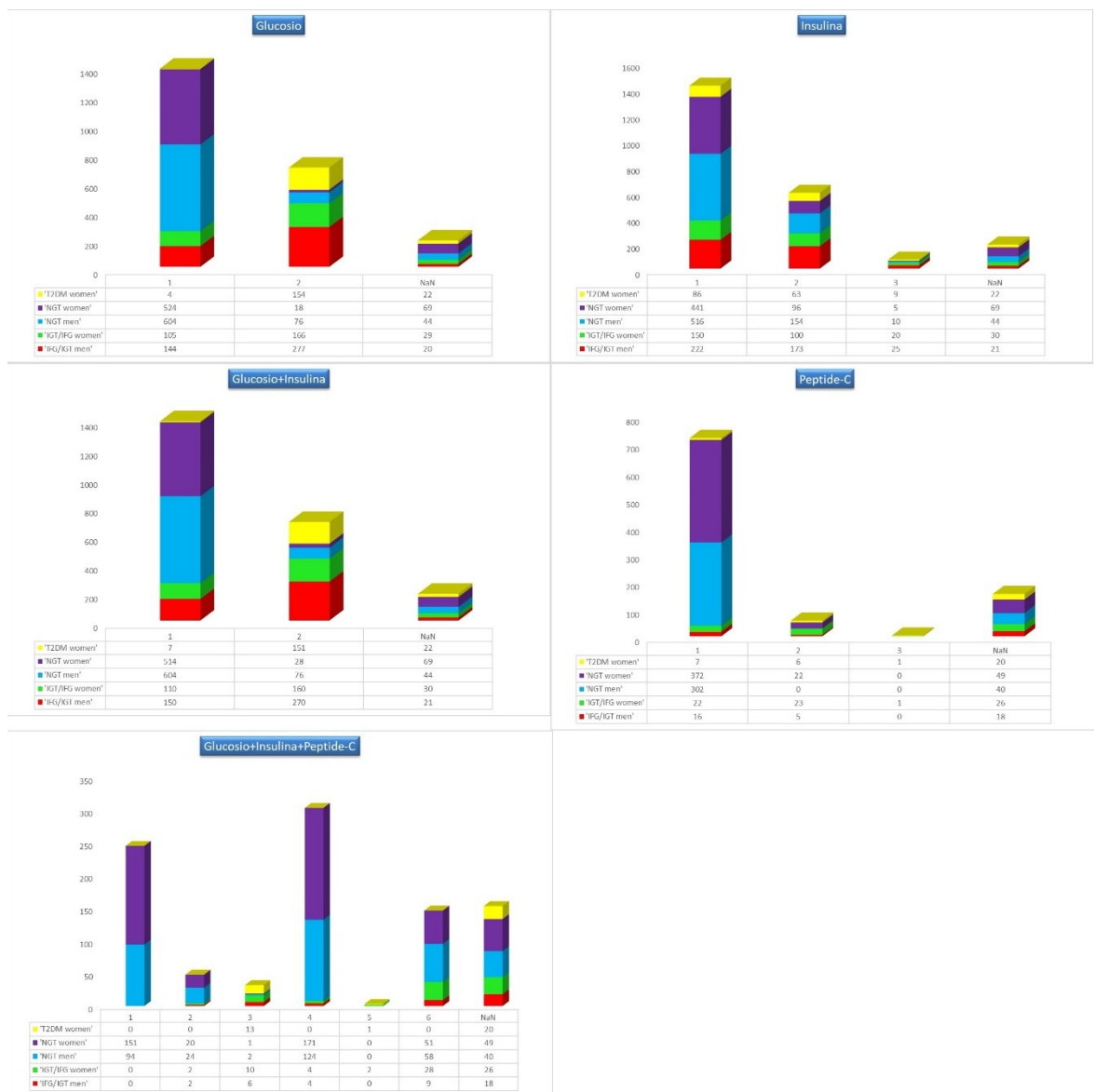
Successivamente, lo studio si è concentrato, attraverso l'utilizzo di grafici excel a colonna, sulla determinazione della tipologia di soggetti (NGT, IGT/IFG, T2DM) appartenenti ad ogni singolo cluster, in base ai 3 metodi ed agli indici della variabile scelta (glucosio, insulina, entrambe, Peptide-C, tutte le variabili) (Figura 8-9-10).



(Figura 8: la popolazione è suddivisa in cluster per ogni variabile a disposizione secondo il metodo Calinski, evidenziando in ogni sottogruppo da che tipo di soggetti è composto)



(Figura 9: la popolazione è suddivisa in cluster per ogni variabile a disposizione secondo il metodo Silhouette, evidenziando in ogni sottogruppo da che tipo di soggetti è composto)



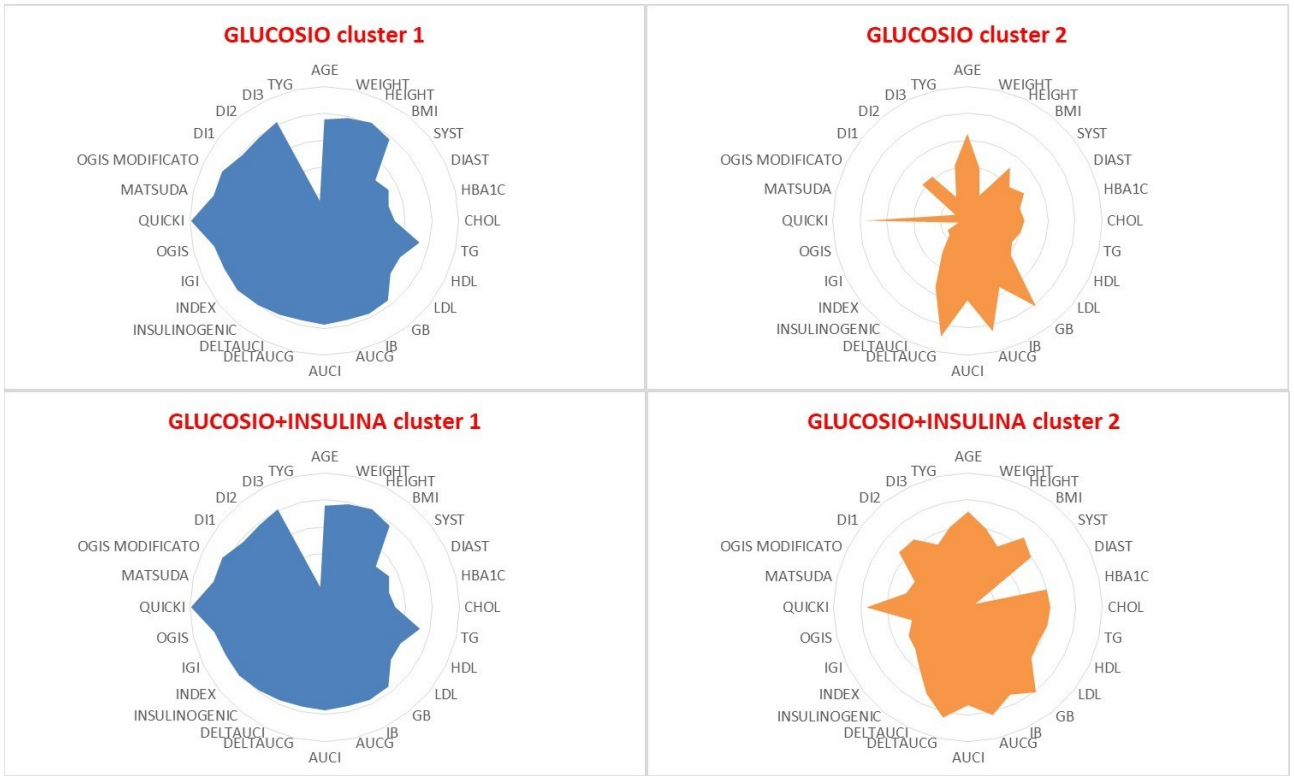
(Figura 10: la popolazione è suddivisa in cluster per ogni variabile a disposizione secondo il metodo Davies, evidenziando in ogni sottogruppo da che tipo di soggetti è composto)

I cluster definiti attraverso il Peptide-C sono stati successivamente giudicati non significativi allo scopo della ricerca, per via della limitatezza dei dati (solo 931 su 2256 soggetti presentavano valori temporali completi per il Peptide-C) e del costo e della difficoltà per ottenerli essendo risultato di test non di routine. Di conseguenza, anche la clusterizzazione effettuato dall'unione di tutte e tre le variabili (glucosio, insulina e Peptide-C) è stata valutata come non significativa. Lo studio ha poi continuato esclusivamente con i cluster ottenuti secondo le tre metodologie in base al glucosio ed a glucosio ed insulina insieme.

Infine i risultati ottenuti dalla clusterizzazione sono stati confrontati con le variabili caratteristiche di ogni soggetto così da poter definire il gruppo patologico o meno caratteristico di ogni cluster (Tabelle 3-4-5) e successivamente i valori di ogni variabile scelta sono stati standardizzati attraverso lo z-score, così da poter evidenziare ulteriormente quali features fossero davvero indicative di ogni cluster (Figura 11-12-13).

(Tabella 3: rappresentano i valori delle variabili caratteristiche dei cluster definiti secondo la metodologia Calinski in base al glucosio ed a glucosio ed insulina insieme)

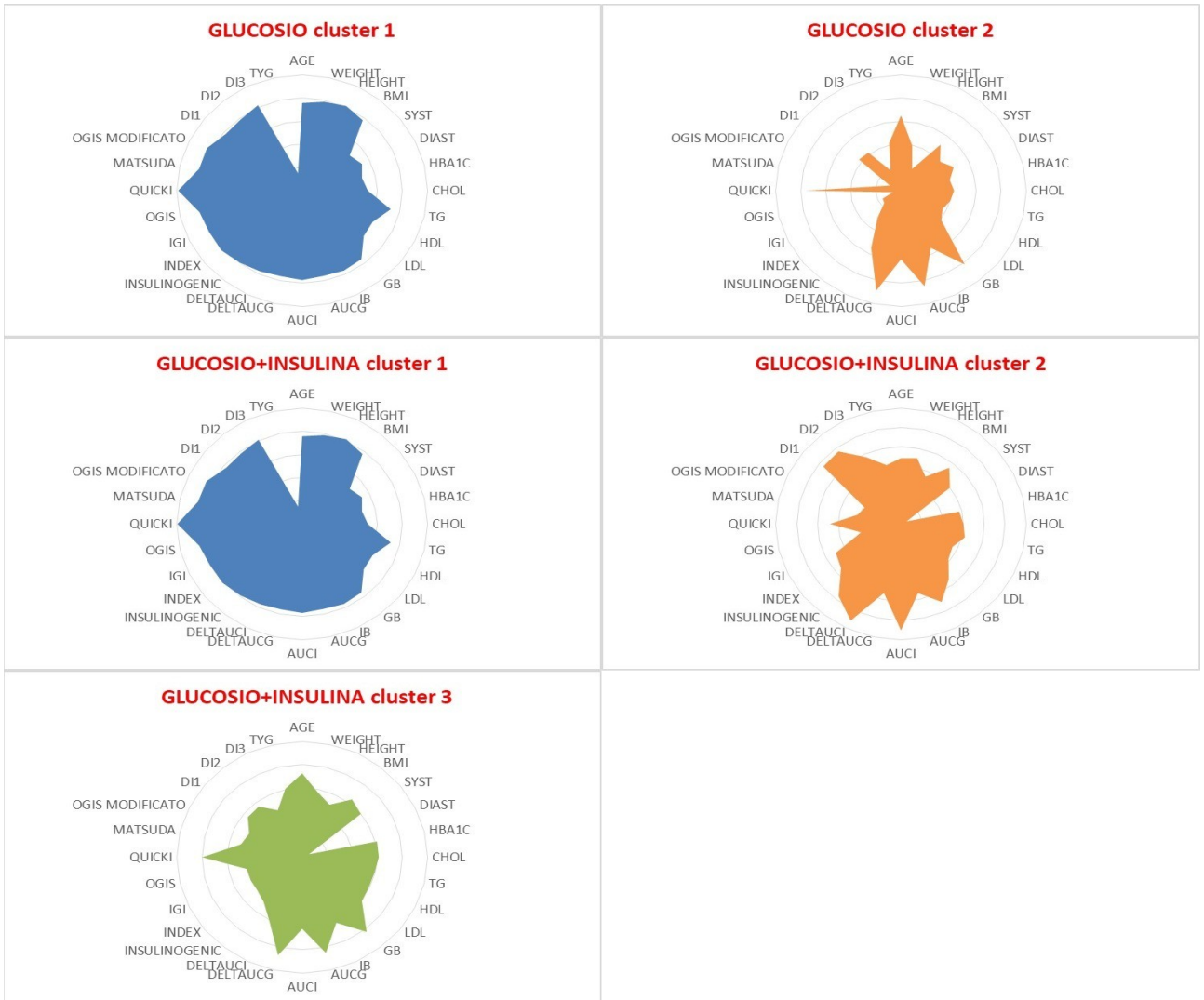
	GLUCOSIO			P-value	GLUCOSIO+INSULINA			
	Cluster 1	Cluster 2			Cluster 1	Cluster 2	P-value	
Age (median(iqr))	39(26,375)	59(14)		0	Age (median(iqr))	39(27)	58(15)	0
Weight (kg) (median(iqr))	73,2(18)	75(17)	0,0025	0	Weight (kg) (median(iqr))	73(17,175)	75(17,5)	<0,001
Height (cm) (median(iqr))	171,1(13,1)	165(12)		0	Height (cm) (median(iqr))	171(13,275)	165(12)	0
BMI (kg/m2) (median(iqr))	24,7(5,1)	27(5,3)		0	BMI (kg/m2) (median(iqr))	24,6(4,9)	27,3(5,1)	0
Systolic blood pressure (mmHG) (median(iqr))	130(25)	140(30)	<0,001	0	Systolic blood pressure (mmHG) (median(iqr))	130(25)	140(30)	<0,001
Diastolic blood pressure (mmHG) (median(iqr))	80(10)	85(10)	<0,001	0	Diastolic blood pressure (mmHG) (median(iqr))	80(10)	85(10)	<0,001
HbA1c (%) (median(iqr))	4,5(0,8)	4,8(0,8)	<0,001	0	HbA1c (%) (median(iqr))	4,5(0,8)	4,8(0,8)	<0,001
Cholesterol total (mg/dl) (median(iqr))	220(65)	235,5(61)	<0,001	0	Cholesterol total (mg/dl) (median(iqr))	220(65)	235(64,5)	<0,001
Triglycerides (mg/dl) (median(iqr))	120(104)	149,5(121)	<0,001	0	Triglycerides (mg/dl) (median(iqr))	117(103)	154,5(125)	<0,001
HDL (mg/dl) (median(iqr))	44(15)	43(15)	NS	0	HDL (mg/dl) (median(iqr))	44(16)	43(14)	0,001318
LDL (mg/dl) (median(iqr))	144,4(49,1)	158,8(59,8)	<0,001	0	LDL (mg/dl) (median(iqr))	145,8(49,55)	158(59,6)	<0,001
Basal Glucose (mg/dl) (median(iqr))	88,2(14,2)	112(18)		0	Basal Glucose (mg/dl) (median(iqr))	88,2(14,4)	112(19)	0
Basal Insulin (µU/ml) (median(iqr))	7(5)	10(6)		0	Basal insulin (µU/ml) (median(iqr))	7(4,6)	11(7)	0
AUCg until 120m (mg/dl) (median(iqr))	14094(4056,75)	21900(4563)		0	AUCg until 120m (mg/dl) (median(iqr))	14100(4074)	21900(4578,75)	0
AUCi until 120m (µU/ml) (median(iqr))	4155(2968,875)	6195(4755)		0	AUCi until 120m (µU/ml) (median(iqr))	4069,5(2663,213)	6630(5253,75)	0
deltAUCg until 120m (mg/dl) (median(iqr))	3456(2757)	8865(3637,5)		0	deltAUCg until 120m (mg/dl) (median(iqr))	3481,5(2772)	8850(3768,75)	0
deltAUCi until 120m (µU/ml) (median(iqr))	3243(2750,25)	4687,5(4095)		0	deltAUCi until 120m (µU/ml) (median(iqr))	3148,5(2401,875)	5145(4647,75)	0
Insulinogenic index (pmol/mmol) (median(iqr))	32,95(21,2825)	28,89(23,275)	<0,001	0	Insulinogenic index (pmol/mmol) (median(iqr))	31,85(19,4125)	31,57(29,0775)	NS
index delta AUC (pmol/mmol) (median(iqr))	120,84(124,095)	56,43(57,38)		0	index delta AUC (pmol/mmol) (median(iqr))	112,655(119,52)	62,96(73,1175)	0
IGI classic (pmol/mmol) (median(iqr))	69,61(66,88)	29,37(31,06)		0	IGI classic (pmol/mmol) (median(iqr))	66,65(66,455)	31,74(36,145)	0
OGIS (ml/ min*m2) (median(iqr))	458,02(87,23)	325,7(128,86)		0	OGIS (ml/ min*m2) (median(iqr))	458,17(85,7575)	323,38(119,3025)	0
QUICKI (w/Log10) (median(iqr))	0,36(0,04)	0,33(0,03)		0	QUICKI (w/Log10) (median(iqr))	0,36(0,04)	0,32(0,03)	0
Matsuda's index (median(iqr))	6,03(4,4375)	2,8(2,0775)		0	Matsuda's index (median(iqr))	6,16(4,2325)	2,62(1,79)	0
OGIS red (ml/ min*m2) (median(iqr))	458,33(86,74)	349,175(69,56)		0	OGIS red (ml/ min*m2) (median(iqr))	459,62(84,21)	345,32(66,17)	0
D11 (OGIS red * AUCi) (median(iqr))	1907547(1201095)	2144383(1409949)	<0,001	0	D11 (OGIS red * AUCi) (median(iqr))	1854568(1114121)	2354453(1532431)	<0,001
D12 (OGIS red * deltAUCi) (median(iqr))	1506069(1077271)	1659876(1255702)	0,001959	0	D12 (OGIS red * deltAUCi) (median(iqr))	1456013(976718,9)	1820535(1346706)	<0,001
D13 (OGIS red * Insulinogenic index)	15284(9348,39)	10309,7(7509,37)		0	D13 (OGIS red * Insulinogenic index)	14621,64(9064,855)	11089,04(8586,03)	0
TYG ((ln(Triglycerides*Basal Glucose))/2) (median(iqr))	4,69(0,45)	4,88(0,4)		0	TYG ((ln(Triglycerides*Basal Glucose))/2) (median(iqr))	4,67(0,43)	4,89(0,38)	0



(Figura 11: rappresentano le variabili significative per ogni cluster ottenuto in base al glucosio ed a glucosio ed insulina insieme secondo il metodo Calinski, i cui valori sono stati standardizzati utilizzando lo z.score, calcolato con le mediane)

(Tabella 4: rappresentano i valori delle variabili caratteristiche dei cluster definiti secondo la metodologia Silhouette in base al glucosio ed a glucosio ed insulina insieme)

	GLUCOSIO			GLUCOSIO+INSULINA						
	Cluster 1	Cluster 2	P-value	Cluster 1	Cluster 2	Cluster 3	P-value (1-2)	P-value (1-3)	P-value (2-3)	
Age (median (iqr))	39(26,375)	59(14)	0	Age (median (iqr))	39(26,5)	55(17)	59(15)	0	0	<0,001
Weight (kg) (median (iqr))	73,2(18)	75(17)	0,002477	Weight (kg) (median (iqr))	72,9(17,375)	82(18)	73(15,7)	<0,001	NS	<0,001
Height (cm) (median (iqr))	171,1(13,1)	165(12)	0	Height (cm) (median (iqr))	171,2(13,1)	167(14)	165(11,25)	<0,001	0	0,036236
BMI (kg/m2) (median (iqr))	24,7(5,1)	27(5,3)	0	BMI (kg/m2) (median (iqr))	24,6(4,675)	29(5,1)	26,7(5,025)	0	<0,001	<0,001
Systolic blood pressure (mmHG) (median (iqr))	130(25)	140(30)	<0,001	Systolic blood pressure (mmHG) (median (iqr))	130(25)	140(27,5)	140(30)	<0,001	<0,001	NS
Diastolic blood pressure (mmHG) (median (iqr))	80(10)	85(10)	<0,001	Diastolic blood pressure (mmHG) (median (iqr))	80(10)	90(15)	85(10)	<0,001	<0,001	NS
HbA1c (%) (median (iqr))	4,5(0,8)	4,8(0,8)	<0,001	HbA1c (%) (median (iqr))	4,5(0,8)	4,65(0,6)	4,8(0,9)	<0,001	<0,001	0,013071
Cholesterol total (mg/dl) (median (iqr))	220(65)	235,5(61)	<0,001	Cholesterol total (mg/dl) (median (iqr))	220(65)	235,5(5,5)	235,5(65)	<0,001	<0,001	NS
Triglycerides (mg/dl) (median (iqr))	120(104)	149,5(121)	<0,001	Triglycerides (mg/dl) (median (iqr))	116(99,5)	181,5(127)	138(117)	<0,001	<0,001	<0,001
HDL (mg/dl) (median (iqr))	44(15)	43(15)	NS	HDL (mg/dl) (median (iqr))	45(16)	40(14)	44(15)	<0,001	NS	0,001428
LDL (mg/dl) (median (iqr))	144,4(49,1)	158,8(59,8)	<0,001	LDL (mg/dl) (median (iqr))	145,6(49,75)	153,7(48,9)	159,5(62,4)	0,037829	<0,001	NS
Basal Glucose (mg/dl) (median (iqr))	88,2(14,2)	112(18)	0	Basal Glucose (mg/dl) (median (iqr))	88,2(15)	106(20)	113,4(18,95)	0	0	<0,001
Basal Insulin (µU/ml) (median (iqr))	7(5)	10(6)	0	Basal Insulin (µU/ml) (median (iqr))	7(4,18)	14,8(10)	9(6)	0	0	<0,001
AUCg until 120m (mg/dl) (median (iqr))	14094(4056,75)	21901(4563)	0	AUCg until 120m (mg/dl) (median (iqr))	13950(4050)	19710(4038)	22530(4756,5)	0	0	<0,001
AUC until 120m (µU/ml) (median (iqr))	4155(296,875)	6195(4755)	0	AUC until 120m (µU/ml) (median (iqr))	4051(2565,375)	11892,5(3885)	5130(2990,25)	0	<0,001	0
deltaAUCg until 120m (mg/dl) (median (iqr))	3456(2757)	8865(3637,5)	0	deltaAUCg until 120m (mg/dl) (median (iqr))	3403,5(2781)	7065(3465)	9300(3877,5)	0	0	<0,001
deltaAUC until 120m (µU/ml) (median (iqr))	3243(2750,25)	4687,5(4095)	0	deltaAUC until 120m (µU/ml) (median (iqr))	3135(2316,375)	9765(3375)	3915(2499,375)	0	<0,001	0
Insulinogenic index (pmol/mmol) (median (iqr))	32,95(21,2825)	28,89(23,275)	<0,001	Insulinogenic index (pmol/mmol) (median (iqr))	31,81(18,5175)	63,78(26,59)	24,22(15,12)	0	0	0
Index delta AUC (pmol/mmol) (median (iqr))	120,84(124,095)	56,43(57,38)	0	Index delta AUC (pmol/mmol) (median (iqr))	111,765(114,335)	151,02(98,36)	46,49(34,925)	<0,001	0	0
IGI classic (pmol/mmol) (median (iqr))	69,61(66,88)	29,37(31,06)	0	IGI classic (pmol/mmol) (median (iqr))	66,365(65,16)	79,38(65,52)	23,85(20,875)	<0,001	0	0
OGIS (ml/min*m2) (median (iqr))	458,02(87,23)	325,7(128,86)	0	OGIS (ml/min*m2) (median (iqr))	461,58(85,82)	303,96(137,805)	337,19(106,1875)	0	0	NS
QUICKI (w/Log10) (median (iqr))	0,36(0,04)	0,33(0,03)	0	QUICKI (w/Log10) (median (iqr))	0,36(0,04)	0,31(0,03)	0,33(0,0325)	0	0	<0,001
Matsuda's index (median (iqr))	6,03(4,4375)	2,8(2,0775)	0	Matsuda's index (median (iqr))	6,34(4,19)	1,89(0,92)	3,08(1,91)	0	0	<0,001
OGIS red (ml/min*m2) (median (iqr))	458,33(86,74)	345,11(69,36)	0	OGIS red (ml/min*m2) (median (iqr))	462,06(82,34)	320,885(66,2)	338,82(61,44)	0	0	<0,001
DI1 (OGIS red * AUCI) (median (iqr))	190747(1201095)	2144383(1409949)	<0,001	DI1 (OGIS red * AUCI) (median (iqr))	1833882(1079099)	3801226(1161367)	1837683(1009529)	0	NS	0
DI2 (OGIS red * deltaAUC) (median (iqr))	1506069(1077271)	1659876(1255702)	0,001959	DI2 (OGIS red * deltaAUC) (median (iqr))	1456013(939717,2)	3173112(1024842)	1444268(832987,5)	0	NS	0
DI3 (OGIS red * Insulinogenic index)	15284(9348,39)	10309,07(7509,37)	0	DI3 (OGIS red * Insulinogenic index)	14489,6(8519,75)	19965,19(8043,39)	8636,98(4999,655)	0	0	0
TYG ((ln(Triglycerides*Basal Glucose))/2) (median (iqr))	4,69(0,45)	4,88(0,4)	0	TYG ((ln(Triglycerides*Basal Glucose))/2) (median (iqr))	4,67(0,42)	4,95(0,335)	4,84(0,41)	0	<0,001	0,0023



(Figura 12: rappresentano le variabili significative per ogni cluster ottenuto in base al glucosio ed a glucosio e insulina insieme secondo il metodo Silhouette, i cui valori sono stati standardizzati utilizzando lo z.score, calcolato con le mediane)

II. Discussione e Conclusione

Lo scopo dello studio si orientava sulla possibilità di stabilire una classificazione fra soggetti NGT, IFG/IGT e T2DM partendo non dalle variabili misurate su ogni soggetto ma utilizzando le curve temporali date dai valori di glucosio (da 0 a 120 minuti), dell'insulina (da 0 a 120 minuti) e del Peptide-C (da 0 a 180). Sono stati ritenuti significativi gli indici di cluster ottenuti attraverso la curve temporali del glucosio e del glucosio ed insulina insieme. Per tutti e tre i metodi (Calinski, Silhouette e Davies), i soggetti in base alla curva temporale glicemica sono stati suddivisi in 2 cluster: il primo è caratterizzato da un'elevata sensibilità all'insulina, in particolare il valore del Quicki (z-score =1,9), ed è costituito per lo più da NGT, mentre il secondo presenta alti valori delle variabili glicemiche (nello specifico, glucosio basale con z-score=1,02, l'AUCg con z-score=1,11 e il deltAUCg con z-score=1,21) ed è formato principalmente da IGT/IFG e T2DM. Nella clusterizzazione definita in base alle curve temporali di glucosio ed insulina prese insieme, si notano delle differenze fra i tre metodi:

- Il metodo Calinski definisce 2 cluster con caratteristiche simili a quelle definite per il glucosio: il primo con elevata sensibilità all'insulina ed il secondo con alti valori delle variabili glicemiche;
- Il metodo Silhouette presenta 3 cluster: il primo uguale a quello già definito per il glucosio (costituito principalmente da NGT), il secondo con elevata secrezione dell'insulina (nello specifico, AUCi con z-score=2,51 e deltAUCi con z-score=2,56) con soggetti di categorie miste ed il terzo con alti valori delle variabili glicemiche e formato principalmente da IGT/IFG e T2DM;
- Il metodo Davies definisce 2 cluster anche in questo caso con caratteristiche simili a quelle dei sottogruppi definiti per il glucosio.

In conclusione, lo studio appena definito risulta utile per allontanarci dal paradigma principale di classificazione dei soggetti in NGT, IFG e/o IGT e T2DM fin'ora utilizzato: infatti, sulla base delle sole curve temporali (principalmente legate al glucosio) permette di definire principalmente 2 grandi cluster con shape completamente diverse, rappresentanti anche soggetto di categorie estremamente diverse (ci sono NGT che hanno una shape molto simile a quella dei soggetti T2DM).

III. Bibliografia

- [1] Howland, Richard D., and Mary Julia Mycek. *Le basi della farmacologia*. Zanichelli, 2007.
- [2] Pasquinelli, F. "Diagnostica e tecniche di laboratorio, vol. 3." *Rosini Editrice srl, Firenze* (1992).
- [3] Russo, Giuseppina T., et al. "Differenze di genere nel diabete mellito di tipo 1 e di tipo 2: i dati italiani." *Epigenetica e malattia metabolica*: 179.
- [4] Casella, Cesare, and Vanni Taglietti. "Principi di fisiologia, Vol. 2." *La Goliardica Pavese Srl* (2002): 249-251.
- [5] Fiorenzo, Conti, et al. "Fisiologia Medica 2." (2005): 719-725. *dica 2. Edi.Ermes s.r.l.* 2010.
- [6] Brunetti, P. *Fisiopatologia, diagnosi e terapia del prediabete*.
- [7] Nuha A. ElSayed, Grazia Aleppo, Vanita R. Aroda, Raveendhara R. Bannuru, Florence M. Brown, Dennis Bruemmer, Billy S. Collins, Marisa E. Hilliard, Diana Isaacs, Eric L. Johnson, Scott Kahan, Kamlesh Khunti, Jose Leon, Sarah K. Lyons, Mary Lou Perry, Priya Prahalad, Richard E. Pratley, Jane Jeffrie Seley, Robert C. Stanton, Robert A. Gabbay; per conto dell'American Diabetes Association, 2. *Classificazione e diagnosi del diabete: standard di cura nel diabete—2023*. *Cura del diabete* 1 gennaio 2023; 46 (Supplemento_1): S19–S40.
- [8] Passamonti, Marco, Gerardo Medea, and Umberto Alecci. "Aggiornamento dei criteri diagnostici per il diabete: l'emoglobina glicata."
- [9] Carlsson, Sofia. "Eziologia e patogenesi del diabete autoimmune latente negli adulti (LADA) rispetto al diabete di tipo 2". *Frontiere in fisiologia* (2019): 320.
- [10] Colpani, Ornella. "MACHINE LEARNING: LA CAPACITÀ DI PREVEDERE APPLICATA ALLA RICERCA E ALLA PRATICA CLINICA *Machine learning: la capacità di previsione applicata alla ricerca e alla pratica clinica.*" *Giornale Italiano di Farmacoeconomia e Farmacoutilizzazione* 11.4 (2019): 5-11.
- [11] Wagner, Robert, et al. "Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes." *Nature medicine* 27.1 (2021): 49-57.

[12] Obura, M., et al. "*Profili clinici dei sottogruppi di glucosio post-carico e la loro associazione con i tratti glicemici nel tempo: uno studio IMI-DIRECT*". *Medicina diabetica* 38.2 (2021): e14428.

[13] Calinski, T., and J. Harabasz. "A dendrite method for cluster analysis." *Communications in Statistics*. Vol. 3, No. 1, 1974, pp. 1–27.

[14] Davies, D. L., and D. W. Bouldin. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. PAMI-1, No. 2, 1979, pp. 224–227.

[15] Kaufman L., and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc., 1990.