

Anders Larsen

Maskinlæring i profesjonell fotball

Kan man ved hjelp av multidimensjonale modeller predikere ikke-kontakt skader i et norsk eliteserielag i fotball?

Masteroppgave i idrettsvitenskap
Institutt for Idrettsmedisinske fag
Norges idrettshøgskole, 2020

Sammendrag

Bakgrunn: Skader i fotball er ansett å være en stor byrde for klubber og spillere, og kan ha en negativ påvirkning på prestasjon og langsiktig utvikling. Det har blitt diskutert at ved å implementere en mer kompleks tilnærming til idrettsskadeforskning, vil det bidra til å forstå og forhindre skader. De fleste eksisterende studier er mono-dimensjonale, som vil si de bare bruker en variabel for å estimere skaderisiko. Predikere skader i fotball er utfordrende da skadene ikke oppstår i en lineær kombinasjon av ulike faktorer. Årsakene til fotballskader er multifaktorale og komplekse. For å kunne håndtere denne kompleksiteten trengs det en multidimensjonal fremgangsmåte som kan forstå og forhindre skade.

Metode: Prosjektet er en prospektiv kohort studie med 32 (18 til 34 år) mannlige fotballspillere fra eliteserien i Norge. Det ble samlet inn data som representerer playerload, acceleration load, total distanse, V4, V5 og høyhastighetsløp gjennom hele 2019 sesongen ved bruk av Catapult OptimEyeX4. Basert på belastningsdataen prøvde vi å predikere ikke-kontakt skade med hjelp av maskinlæringsmodeller.

Resultater: Det ble registrert 34 ikke-kontakt skader gjennom sesongen. Ved bruk av innhøstede data, klarte maskinlæringsmodellene Support vector machine, Decision tree, Random forest og Logistisk regresjon predikerte 0 % av skadene.

Konklusjon: Ingen av maskinlæringsmodellene klarte å predikere ikke-kontaktskader i fotball basert på datagrunnlaget en sesong. GPS-data i seg selv ikke er nok for å kunne predikere et så sammensatt problem.

Nøkkelord: Maskinlæring, prediksjon, GPS-data, treningsbelastning, ikke-kontaktskader

Innhold

Sammendrag	3
Forkortelser og begrepsavklaring	6
Forord	8
1. Introduksjon	9
1.1 Formål med studien	10
1.1.1 Problemstilling.....	10
2. Teori	11
2.1 Skade - sykdomsforekomst i fotball	12
2.2 Risikofaktorer og skademekanismer i fotball	13
2.3 Treningsbelastning.....	14
2.4 Monitorering av treningsbelastning – hvorfor og hvordan?	18
2.4.1 Validitet og reliabilitet av GPS-enheter.....	21
2.5 Maskinlæring	21
2.6 Belastningsstyringszyklusen	26
3. Metode	29
3.1 Studiedesign.....	29
3.2 Utvalg	29
3.3 Datainnsamling	29
3.4 Skadedefinisjon	30
3.4.1 Skadekategorier	30
3.5 Utstyr og målemetode	31
3.6 Dataeksport	31
3.7 Catapult OptimEye X4.....	32
3.8 Player Load™	32
3.9 Datasett	33
3.10 Maskinlæringsmodellene.....	33
3.11 Valg av parameter	34
3.12 Parameter	34

3.12.1	Mean Standard deviation Workload Ratio (MSWR).....	34
3.12.2	Exponential Weighted Moving Average (EWMA).....	35
3.12.3	Skadeparameter.....	35
3.12.4	Presisjon.....	36
3.12.5	Sensitivitet	36
3.12.6	F1 score.....	37
3.12.7	Areal under kurven (AUK).....	37
3.12.8	Adaptere syntetisk data.....	37
3.13	Statistikk	37
3.14	Etikk.....	38
3.15	Budsjett og samarbeid	38
4.	Resultater.....	39
4.1	Skader	40
5.	Diskusjon	47
5.1	Prediksjon av skade i fotball.....	47
5.2	Er maskinl�ring basert p� GPS-data komplekst nok?	50
5.3	Assosiasjon og prediksjon	51
5.4	Skadelokalisasjon og skademekanismer	52
5.5	Valg av skadedefinisjon.....	52
5.5.1	Valg av monitoreringsmetode.....	54
5.6	Begrensinger i denne studien	55
5.7	Videre forskning	57
6.	Konklusjon.....	58
	Referanser.....	59
	Tabelloversikt	76
	Figuroversikt.....	77
	Vedlegg	78

Forkortelser og begrepsavklaring

AC	Acceleration load
ADASYN	Det er en metode som brukes for å korrigere ubalanse i et datasett
AUK	Arealet under kurven
DT	Decision tree
EWMA	Exponential weighted moving average
F1	F1-score
GPS	Global posisjon system
HHL	Høyhastighetsløp. Distanse løpt over 20 km/t
IK	Ikke-kontakt skade
IK-F	Ikke-kontakt skade som fører til fravær
KD-2	Kampdag – 2 (To dager før kamp)
KD-3	Kampdag – 3 (Tre dager før kamp)
KD-4	Kampdag – 4 (Fire dager før kamp)
KI	Kunstig intelligens
LG	Logistisk regresjon

MSWR	Mean standard deviation workload
Overfitting	Når treningsdata og testdata er for lik. Så kan det oppstå overfitting fordi modellen lærer detaljene i treningsdataen, og bruker det den har lært til å predikere på testdataen.
PL	Player Load™
RF	Random Forest
Sensitivitet	Modellens evne til å identifisere type skade
STD	Standardavvik
SVM	Support Vector Machine
Syntetisk data	Data som ikke er virkelig, eller som er «funnet opp» av maskinlæringsmodellen
TD	Totale distanse
V4	Totale distansen løpt mellom 20-24,9 km/t
V5	Totale distansen løpt over 25 km/t

Forord

Da var tiden kommet for å endelig levere inn masteroppgaven min - det har vært noen fantastiske år på NIH! Jeg sitter igjen med kunnskap, venner og minner for livet.

Først må jeg takke mine to veiledere Thor Einar Andersen og Torstein Dalen-Lorentsen for all hjelp med oppgaven. En spesiell takk til Torstein som lot meg være med i Strømsgodset, en god erfaring jeg ikke ville vært for uten. Mitt beste minne fra datainnsamlingen er når Torstein dro en måned til Australia, og jeg fikk muligheten til å være vikar som fysisk trener for klubben.

Tusen takk til Strømsgodset IF og alle folkene i og rundt klubben. Det var fantastisk gøy å bli kjent med og få være en del av teamet en hel sesong.

Jeg vil også rekke en stor takk Garth Theron som lagde maskinlæringsmodellen. Tusen takk for samarbeidet og all hjelp.

At jeg skulle avslutte mitt siste år som student midt i koronakrisen er en spesiell følelse. Samtidig var jeg heldig at prosjektet mitt ikke ble påvirket og jeg fikk skrevet ferdig oppgaven og levert i tide.

Avslutningsvis må jeg takke Rune Engen og Jenny Aambø for godt samarbeid gjennom to år på masterstudie. Takk til far og søster for korrekturlesning. Og Ida, for at du har latt meg bruke 90 % av kjøkkenbordet de siste par månedene.

Anders Larsen

Oslo, Juni 2020

1. Introduksjon

Idrett har utviklet seg fra å være en lek, et spill for å underholde og for å være en fritidsaktivitet, til å bli en profesjonell industri. Idrettsutøvere skal imøtekomme krav fra sponsorer og en tett pakket kalender med kamper, turneringer og andre arrangement (Soligard et al., 2016). Stor ukentlig treningsmengde og et tett kampprogram i konkurransesesongen gjør det vanskelig for utøvere å absorbere den totale belastningen de blir påført (Schwellnus et al., 2016; Soligard et al., 2016). Det kan resultere i dårligere prestasjon og er også assosiert med en økt risiko for skade (Soligard et al., 2016). Forskningen er tvetydig og Gabbett, Hulin, Blanch, og Whiteley (2016) fant i sin studie at høy treningsbelastning over tid kunne ha en beskyttende effekt på skaderisikoen, men at det var store endringer i belastning som ga økt risiko for skade. Å kunne forutse (predikere) skade er vanskelig, men samtidig en nøkkelfaktor for å forhindre skade (Bittencourt et al., 2016). Skader hos fotballspillere er vanlig, i gjennomsnitt må hver spiller forvente rundt to skader per sesong. Dette tilsvarer ca. 50 skader per sesong for et fotballag med en spillerstall på 25 spillere (Ekstrand, Hägglund, & Waldén, 2011b). Skader hos fotballspillere har en stor bivirkning både på prestasjon og utvikling til spilleren, og kan være hemmende for langsiktig spillerutvikling (Gabbett et al., 2016), i tillegg ha en negativ effekt på lagets prestasjon (Hägglund et al., 2013b). Skader representerer også en økonomisk byrde for profesjonelle klubber. Klubbene bruker mye penger på spillere som er skadet, kostnadene er knyttet til spillerens rehabiliteringsperiode og medisinsk behandling. I tillegg taper klubben penger på at spilleren ikke spiller kamper, og dermed ikke bidrar til å tjene inn penger til klubben (Hägglund et al., 2013b). Forskning som ble gjort i Spansk 1- og 2. divisjon viser at skader gir ca. 16% fravær fra trening og kamp hos en profesjonell fotballspiller, og dette medfører en betydelig kostnad for lagene (Fernández Cuevas et al., 2010).

Studier av skader hos fotballspillere har manglet pålitelig data som beskriver spillernes treningsbelastning. I nyere tid har det blitt gjennomført studier med mål om å finne gode tiltak som kan redusere skader på spillerne. Ved gjennomføring av disse studiene har det vært brukt Global Posisjonssystem (GPS) (Ehrmann, Duncan, Sindhusake, Franzsen, & Greene, 2016; Rago et al., 2019). Ved bruk av GPS kan vi hente gode treningsbelastningsdata fra hver trening og kamp (Alessio Rossi, 2018). Disse dataene

viser i detalj en spillers bevegelse på banen, og kan brukes til å identifisere treningsmønstre til spilleren og bidra til taktisk analyse.

De fleste eksisterende studier av treningsbelastning er mono-dimensjonale, der de har brukt kun en variabel om gangen for å estimere skaderisiko (Arnason, Andersen, Holme, Engebretsen, & Bahr, 2008; Petersen, Thorborg, Nielsen, Budtz-Jørgensen, & Hölmich, 2011). Utdfordringen med å predikere skade i fotball er at skadene ikke oppstår fra en lineær kombinasjon av ulike faktorer. Årsakene til fotballskader er multifaktorale og svært komplekse (Bittencourt et al., 2016). De ulike faktorene kan være koblet til hverandre på en ikke-lineær måte, og små endringer i faktorene kan føre til uventede forandringer i de andre faktorene. Manuel Stein (2017) mente i sin studie at vi ikke har klart å utnytte de komplekse mønstrene som ligger i innsamlede data, noe som samsvarer med Bittencourt et al. (2016). For å klare og avdekke den komplekse årsaksforklaringen til fotballskader, trengs det en kompleks tilnærming og forklaringsmodell (Bittencourt et al., 2016). Klubber på elitenivå er interesserte i praktiske og nyttige verktøy/hjelpemidler som kan bidra til at det medisinske støtteapparat, trenerteamet og ikke minst spillerne kan ta beslutninger for å kunne senke risikoen for en ny skade kan oppstå (Kirkendall & Dvorak, 2010).

1.1 Formål med studien

Hva som forårsaker skade i fotball er sammensatt. I søken etter å forstå denne kompleksiteten er det vanlig å bruke en multi-faktoriell modell for å kunne avdekke skjulte sammenhenger og trender (Bittencourt et al., 2016). Komplekse sammenhenger krever bruk av multi-dimensjonale modeller. Formålet med studien er å predikere ikke-kontakt skade basert på spillernes belastningsdata. Det er brukt en eksisterende maskinlæringsmodell fra Alessio Rossi (2018), og sett om vi klarer å bruke den samme modellen på vårt datasett. Videre håper vi dette kan gi oss økt innsikt i hvordan vi kan planlegge trening bedre for å unngå skader.

1.1.1 Problemstilling

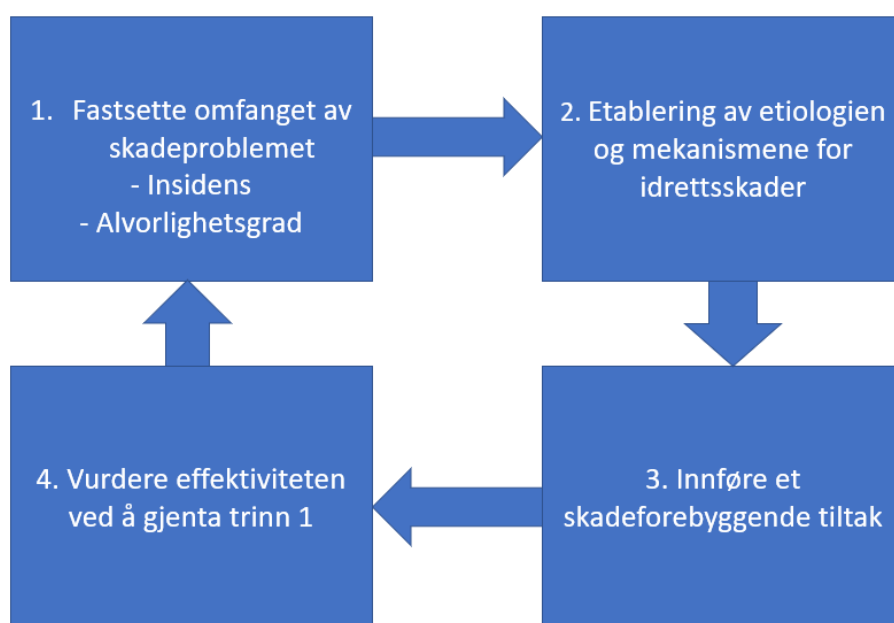
Kan man ved hjelp av multidimensjonale modeller predikere ikke-kontakt skade i et norsk eliteserielag i fotball?

2. Teori

Fotball er blant verdens største idretter, på verdensbasis er det estimert at fotball spilles av 270 millioner mennesker, av disse er 38 millioner registrert i klubber, og mer enn 100 000 av disse er profesjonelle spillere (FIFA, 2007). Fotball er den mest populære idretten i Norge, og Norges fotballforbund (NFF) har mer en 375 000 registrerte spillere (Norges Idrettsforbund, 2014).

Skader er vanlig i lagidretter, og kan ha økonomiske, fysiske og psykiske konsekvenser for både spillerne og laget (Ruddy et al., 2019). Mange av spillerne opplever skader med påfølgende smerter, uførhet og økonomiske belastninger knyttet til skaden (Petersen et al., 2011).

Verdens helseorganisasjon startet et prosjekt på 1980-tallet som fikk navnet Sport for all: Sports injuries and their Prevention. Dette prosjektet skulle bidra til en økt forståelse av bakenforliggende årsaker til idrettsskader, med det formål å kunne videreutvikle skadeforebyggende tiltak (van Mechelen, 1997). I dette prosjektet definerte de fire hovedpunkter som skulle innhente informasjon for å øke forståelsen for idrettsskader (Figur 1).



Figur 1: Stegene for utviklingen av skadeforebyggende tiltak (van Mechelen, 1992)

Steg 1 og 2 er de epidemiologiske fasene som kartlegger omfanget, patogenesen og etiologien av idrettsskadene. Det første steget i modellen er å kartlegge å beskrive

skadeomfang, hyppighet og alvorlighetsgrad av skadene. Denne oppgaven vil gå under steg 2 og ved bruk av «maskinlæring» som metode, skal studien søke å finne skademekanismer, risikofaktorer og årsakssammenhenger.

I steg 3 og 4 skal den nye kunnskapen anvendes for å utvikle, implementere og evaluere skadeforebyggende tiltak. Basert på kunnskapen fra steg 1 og 2 kan en intervensjon som har til hensikt å redusere skadeomfanget og alvorlighetsgradene av skadene utvikles og implementeres (van Mechelen, Hlobil, & Kemper, 1992).

2.1 Skade - sykdomsforekomst i fotball

Fotball er en kompleks idrett med forholdsvis høy risiko for å utvikle eller bli skadet (Pfirschmann, Herbst, Ingelfinger, Simon, & Tug, 2016). De vanligste skadene er muskelskader i hamstring, quadriceps og lysken, samt ligamentskader i kne- og ankelregionen (Ekstrand et al., 2011b). Studier som ser på skaderisiko hos fotballspillere tar ofte bare for seg en sesong eller en turnering (Andersen, Tenga, Engebretsen, & Bahr, 2004; Arnason et al., 2004; Ekstrand, Waldén, & Häggglund, 2004; Waldén, Häggglund, & Ekstrand, 2005a; Waldén, Häggglund, & Ekstrand, 2005b). Noen få studier har hentet data fra to eller flere sesonger (Arnason et al., 2008; Häggglund, Waldén, & Ekstrand, 2006). Dette gjør at det har vært vanskelig å analysere sammenhenger og skadevariasjonene mellom to og flere sesonger.

Det er funnet at skadeinsidensen hos profesjonelle fotballspillere varierer mellom 2,48 til 9,4 skader per 1000 time eksponering (Ekstrand, Häggglund, & Waldén, 2011a; Pfirschmann et al., 2016; Waldén et al., 2005b). Skadeinsidensen under kamp viser seg å være høyere, med en skala mellom 15 til 20 skader per 1000 timer med kamp (Faude, Röbler, & Junge, 2013). Å konkludere ut fra disse tallene er utfordrende da det er benyttet ulike definisjoner av skade i de forskjellige artiklene.

Ekstrand og hans medarbeidere (2011) fulgte 23 profesjonelle lag gjennom 7 sesonger. De registrerte 4483 skader løpet av perioden. Tatt utgangspunkt i denne studien så vil det i snitt si at en spiller vil ha 2 skader pr sesong, noe som resulterer i 50 skader hos et lag med 25 spillere (Ekstrand et al., 2011b). Av disse skadene oppsto 57% under kamp, mens 43 % oppsto under trening. Overbelastning stod for 28% av skadene. I gjennomsnitt mistet hver spiller 37 dager på grunn av skader, som betyr at spillerne mister ca. 12 % av sesongen (Ekstrand et al., 2011b). Insidenstallene i studien til

Ekstrand og hans medarbeidere (2011) er konsistent med tall fra den svenske og norske eliteserien i fotball (Andersen et al., 2004; Hägglund et al., 2006; Waldén et al., 2005b).

2.2 Risikofaktorer og skademekanismer i fotball

For å kunne utvikle forebyggende programmer og modeller må en vite hvilke risikofaktorer som er forbundet med skade hos fotballspillere (Hägglund, Waldén, & Ekstrand, 2013a; van Mechelen et al., 1992). Risikofaktorer varierer mellom individer innenfor samme idrett, noe som gjør skadeprediksjon til en av de mest utfordrende problemstillingene i idrettsskadeforskning (Bittencourt et al., 2016).

Det har både blitt forsket på personrelaterte risikofaktorer og miljørelaterte risikofaktorer for skader. Personrelaterte faktorer som tidligere skade, høy alder, tretthet, dårlig bevegelse, lav muskelstyrke eller ubalanse i styrken mellom forside og bakside lår har blitt indentifisert som risikofaktorer for skade i underekstremiteten (Arnason et al., 2004; Engebretsen, Myklebust, Holme, Engebretsen, & Bahr, 2010; Fousekis, Tsepis, Poulmedis, Athanasopoulos, & Vagenas, 2011; Henderson, Barnes, & Portas, 2010; Hägglund et al., 2006). Risikofaktorene knyttet til miljøet spillerne spiller og trener i, har ikke blitt undersøkt i like stor grad, men å spille kamper er assosiert med en økt skaderatio hos spillerne (Ekstrand et al., 2011a; Engebretsen et al., 2010). Utmattelse hos spilleren kan også være en komponent for økt skaderisiko, der noen studier har funnet at muskelskader oppstår oftere mot slutten av kamper (Ekstrand et al., 2011a; Hawkins & Fuller, 1999).

Skader i muskelgruppene hamstring, quadriceps, lyskene og legg står for mer enn 90 % av alle muskelskader i underekstremiteten (Hägglund et al., 2013a). Lyske- og quadricepsskader er mer vanlig i den dominante foten til spilleren, og dette kan begrunnes med et at den dominante foten påføres en større belastning i form av skudd og pasninger. Hamstring- og leggskaader er uavhengig av fot (Hägglund et al., 2013a).

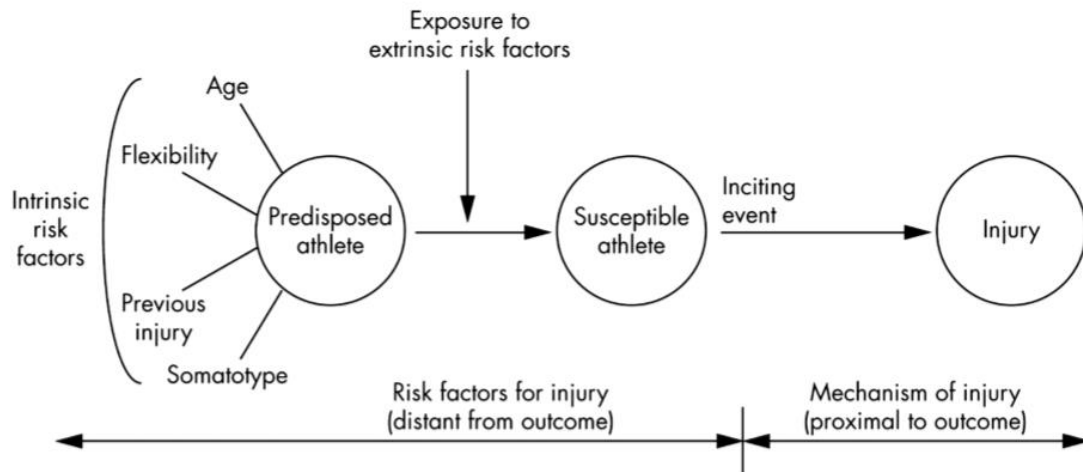
Kamprelaterte risikofaktorer som kampresultat, hjemme eller bortekamp har vist seg å påvirke den generelle skadeprevalensen i fotball (Hägglund, Waldén, & Ekstrand, 2009). Hamstring og lyske skader har en større sannsynlighet for å oppstå under hjemmekamp, i motsetning til bortekamp. Mer tid med ball innad i laget og mer tid brukt i angrep har blitt registrert og dokumentert ved hjemmekamper, og dette kan være

noe av grunnen til at en ser en litt høyere risiko for hamstring og lyskeskader ved hjemmekamper (Hägglund et al., 2013a).

2.3 Treningsbelastning

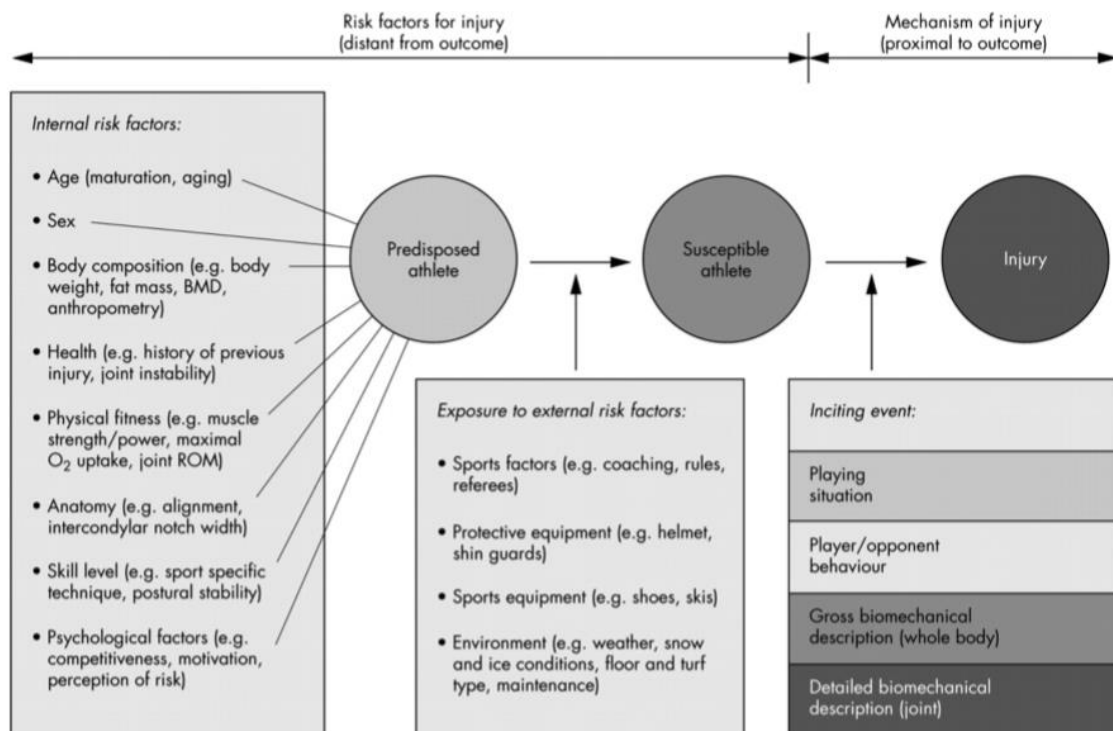
Årsaken til skader refereres til som etiologi. Etiologi omhandler læren om hvordan en sykdom eller skade oppstår (Meeuwisse, 1994). Skader oppstår gjennom komplekse interaksjoner mellom flere risikofaktorer, og den multifaktorelle naturen av idrettsskader er velkjent (Bahr & Holme, 2003; Meeuwisse, Tyreman, Hagel, & Emery, 2007). Ser man på en enkel risikofaktor om gangen klarer en ikke å forutsi skader tilstrekkelig, og begrensningene av monodimensjonale analyser av risikofaktorer er blitt diskutert (Meeuwisse, 1994). Det er anbefalt at en bruker multivariate metoder, som gjør at en kan se på flere risikofaktorer samtidig (Bahr & Holme, 2003). Opp igjennom årene har det blitt presentert flere skadeetiologimodeller som har som formål å forklare årsaken til idrettsskader (Bahr & Krosshaug, 2005; Bittencourt et al., 2016; Meeuwisse, 1994; Meeuwisse et al., 2007; Windt & Gabbett, 2017).

Den første multifaktorelle modellen ble presentert av Meeuwisse og hans medarbeidere i 1994 (se figur 2). Meeuwisse og medarbeidere (1994) brukte kroniske-sykdomsmodeller som grunnlag. De brukte disse modellene til å bygge videre på, for å kunne anbefale epidemiologisk multifaktoriell tilnærming for å forstå sammenhengen som forårsaker skadene (Meeuwisse, 1994). I modellen til Meeuwisse og hans medarbeidere (1994) har utøverne noen interne risikofaktorer som i seg selv er forbundet med skade. Noen av disse faktorene er ikke-modifiserbare faktorer (eks. alder, høyde), mens andre er modifiserbare faktorer (eks. stabilitet). Utøverne blir deretter disponert for eksterne risikofaktorer som spilloverflaten og utstyr, samt motstandere som eksponerer spillerne for skade. Til slutt oppstår en hendelse der den biomekaniske belastningen overstiger vevets toleranse, og det oppstår en skade (Meeuwisse, 1994).



Figur 2: Meeuwisse og hans medarbeidere 1994, første multifaktorelle skademodell.

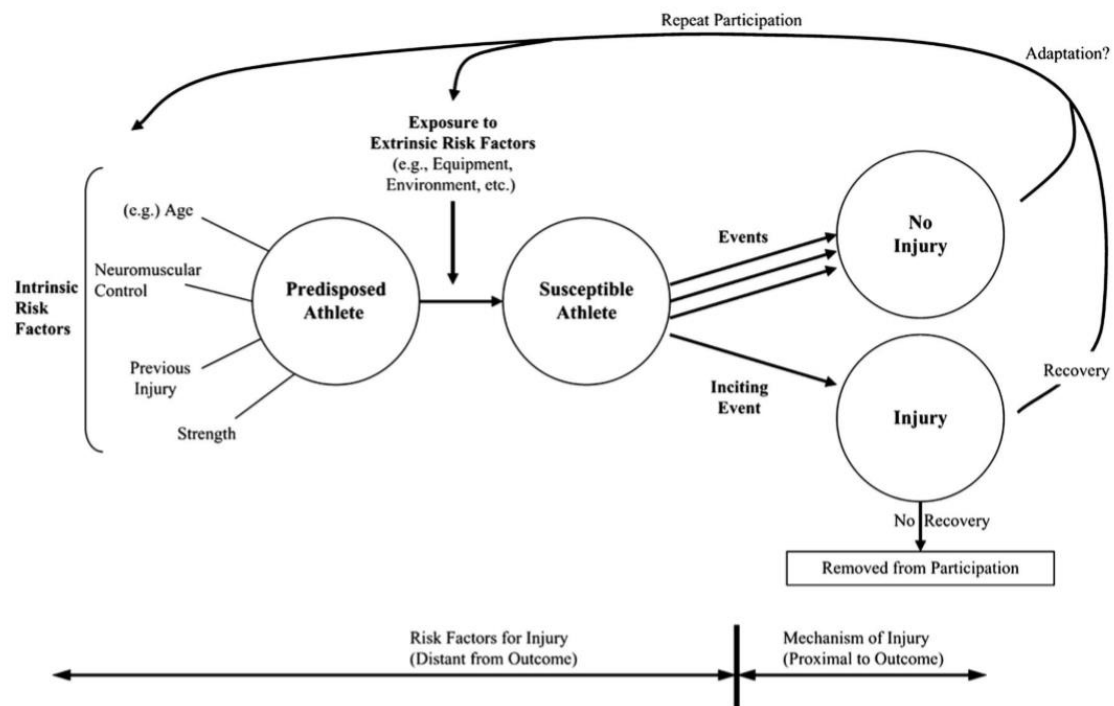
I to påfølgende artikler (Bahr & Holme, 2003; Bahr & Krosshaug, 2005) bygde Bahr og hans medarbeidere videre på den første modellen til Meeuwisse og hans medarbeidere (1994). De utvidet beskrivelsen av interne og eksterne risikofaktorene, og utarbeidet en grundigere beskrivelse av de biomekaniske faktorene som bidrar til skade (se figur 3).



Figur 3: Bahr og Krosshaugs etiologimodell (Bahr & Krosshaug, 2005).

Meeuwisse og hans medarbeidere (2007) kom med en modifisering av etiologimodellene i form av en mer dynamisk modell (figur 4). Denne modellen viser at

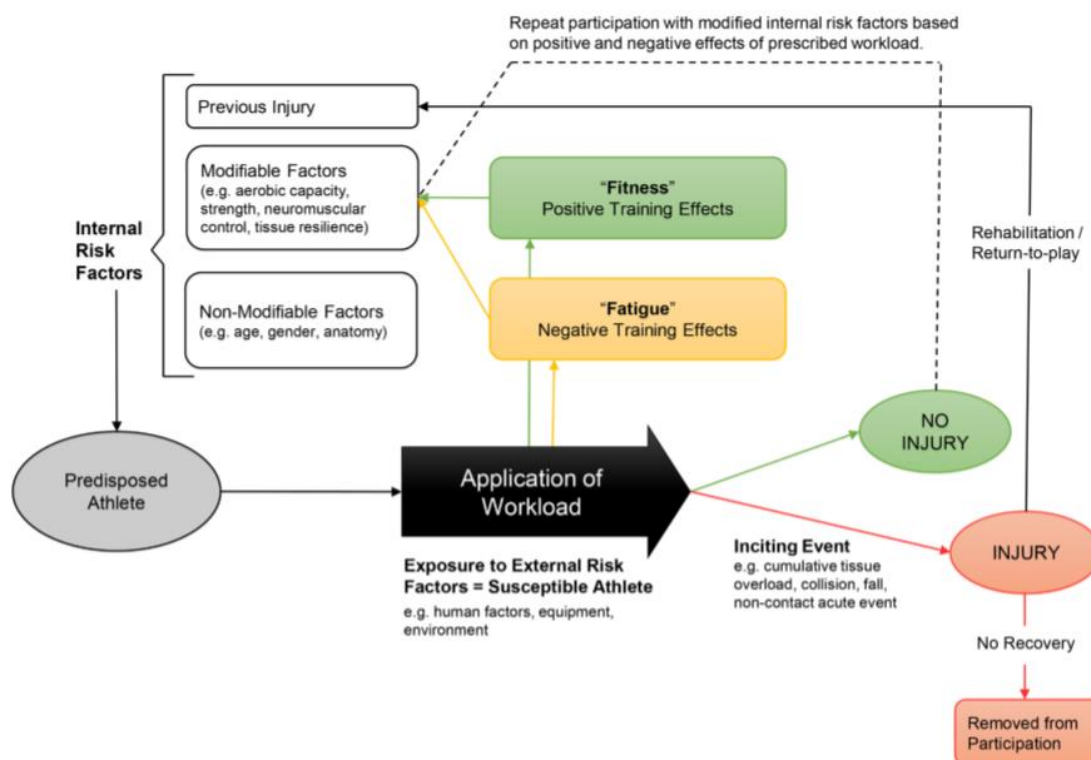
tidligere lineære modeller ikke ivaretar den dynamiske og ikke-lineære karakteren til idrettsutøvers skaderisiko (Meeuwisse et al., 2007). I tilfeller der deltakelsen ikke resulterer i skade, endrer den faktiske risikofaktoren seg gjennom fysiologiske tilpasninger (Windt & Gabbett, 2017). Og i de tilfellene det resulterer i en skade, kan utøveren komme tilbake i spill med en modifisert skaderisiko, og i noen tilfeller aldri komme tilbake til idretten (Windt & Gabbett, 2017).



Figur 4: Meeuwisse og hans medarbeidere (2007) sin dynamiske etiologimodell.

Alle idrettsskader skjer mens utøveren er eksponert for belastning, men de tidligere etiologimodellene har ikke hatt med belastning som en faktor i modellen, eller har forklart den sterke assosiasjonene mellom belastning og skade (Windt & Gabbett, 2017). Windt og Gabbett (2017) kom med en oppdatert modell som inneholder belastning som en faktor. De forklarer at treningsbelastning eller konkurransebelastning i seg selv ikke forårsaker skader. En utøver kan ha en uventet stor belastning, og deretter få en skade i fremre korsbånd, men skaden oppstår fremdeles på grunn av en hendelse (eks. en dynamisk valgus kollaps). Belastningsskader kan være belastningsrelaterte, og det er ikke alltid en klarer å identifisere hendelsen, men skaden har likevel alltid en hendelse av kumulativ vevsbelastning (Clarsen, Myklebust, & Bahr, 2013). Med denne forståelsen er ikke skade direkte forårsaket av belastningen utøveren blir påført. I stedet påfører trening- og konkurransebelastning en skaderisiko

ved å eksponere utøverne for potensielt skadelige situasjoner, samt gjennom positive og negative effekter på modifiserbare interne risikofaktorer (Windt & Gabbett, 2017). Med dette som grunnlag forslår Windt og Bahr (2017) en viderutviklet skadeetiologi-modell som eksplisitt inkorporerer arbeidsmengde i årsakskjeden.



Figur 5: Belastnings-etiologiskademodellen (Windt & Gabbett, 2017).

Windt og Gabbett (2017) har delt interne risikofaktorer inn i modifiserbare og ikke-modifiserbare faktorer. Dette ble gjort for å skille faktorene som kan endres gjennom tilpasning fra faktorene som ikke kan endres. Videre har de utvidet tidligere modeller ved å inkludere «Application of Workload» som den primære prosessen hvor utøveren blir utsatt for forskjellige ytre risikofaktorer og hendelser som potensielt kan forårsake skade.

Hvis en utøver påføres en skade, må utøveren delta i et rehabiliteringsprogram for å klare å returnere til spill igjen. I rehabiliteringsperioden er utøveren avhengig av treningsbelastning for å gjenopprette motstandskraft i det skadde vevet. Og på den måten adaptere til de fysiske kravene som spillere møter i trening og kamp (Blanch & Gabbett, 2016).

Windt og Gabbett (2017) hevder i sin studie at skader i idrett kan skyldes inadekvat belastningsstyring. De hevder at treningsbelastning i hovedsak påvirke skadeetiologi på tre måter; eksponering, fitness og utmattelse. Eksponering blir forklart som trening- og konkurransebelastning som utsetter utøveren for eksterne risikofaktorer, og gjennom disse blir de utsatt for hendelser hvor de kan bli skadet. Fitness begrepet uttrykker positive tilpasninger forbundet med trening, som kan forbedre de interne modifiserbare risikofaktorene. (F.eks. Aerob kapasitet, kroppssammensetning). Når en trener er det er ønskelig å få en positiv treningseffekt, og økt aerob og anaerob kapasitet vil kunne motvirke muskulær utmattelse. I følge Windt og Gabbett (2017) er utmattelse en negativ konsekvens assosiert med treningsbelastning. Ved høy og intensiv trening brytes cellene i kroppen ned, som fører til midlertidig nedsatt fysisk kapasitet (Windt & Gabbett, 2017). Nedsatt fysisk kapasitet kan føre til økt interne risikofaktorer på grunn av at vevstoleransen svekkes og nevro-muskulær kontroll blir dårligere. Dette kan øke risikoen for skade og kroppen bør gjenoppbygges i en restitusjonsprosess (Windt & Gabbett, 2017)

Evidensen viser også at en høy kronisk belastning kan gi en beskyttende effekt mot skade (Gabbett et al., 2016; Hulin et al., 2014). Studier har vist at utøvere som har gode fysiske egenskaper som styrke, hurtighet og grunnlag for flere sprinter har en høyere toleranse for en høy belastning og redusert risiko for skade (Malone et al., 2017b).

2.4 Monitorering av treningsbelastning – hvorfor og hvordan?

Interessen for belastningsstyring har økt betraktelig siden år 2000 (Gabbett, 2018), og en betydelig mengde forskning har sett på forholdet mellom treningsbelastning og skade (Eckard, Padua, Hearn, Pexa, & Frank, 2018). Både for å optimalisere prestasjonsevnen, legge til rette for best mulig fremgang i treningsarbeidet og for å forebygge skader (Cardinale & Varley, 2017; Gabbett, 2018). Å kunne kvantifisere belastning er et nyttig verktøy i forskning, men ikke minst i treningsarbeid, særlig i toppidrett hvor en jobber for å unngå overtrening, skade og sykdom (Cardinale & Varley, 2017). Grunnet den økende interessen rundt temaet treningsbelastning, både hos spillere, trenere, lagledere, samt bekymringen knyttet til dårlig belastningsstyring og utøvernes helse. Presenterte International Olympic Committee (IOC) en konsensusstudie om forholdet mellom belastning og skaderisiko (Soligard et al., 2016). De konkluderte med at det er en

sammenheng mellom inadekvat trenings-, kampbelastning- og skaderisiko innen ulike idretter (Soligard et al., 2016).

Treningsbelastning har blitt registrert på mange måter opp igjennom tidene. Alt fra selvpoplevd belastningsregistrering ved bruk av treningsdagbok til mer teknologiske hjelpemidler som GPS er brukt for å monitorere treningsbelastningen (Ehrmann et al., 2016). Treningsbelastning er definert som den kumulative mengden stress påført en utøver fra en eller flere treningsøkter over en periode (Eckard et al., 2018).

Treningsbelastning kan måles gjennom ulike dimensjoner, det brukes intern/ekstern, subjektiv/objektiv og absolutt/relativ mål på treningsbelastning.

Treningsbelastning blir delt inn i to hovedgrupper, interne- og eksterne belastninger. Ekstern treningsbelastning er for eksempel distanse løpt, akselerasjon, sprintmeter eller kamper spilt (Soligard et al., 2016). Gjennom teknologisk utstyr kan en overvåke den eksterne belastningen som påføres utøveren. Disse tallene kan gi en beskrivelse av den eksterne belastningen til spilleren i en kamp eller under trening. Det blir ansett som et viktig verktøy fordi den tar opp bevegelser og intensitetsendringer spillerne blir utsatt for (Rampinini et al., 2011).

Intern treningsbelastning kan være subjektiv eller objektiv. Subjektive målinger er de som er selvrapporterte. Det kan være «session rate of perceived exertion (sRPE)» som er en CR10-scale, hvor hver enkelt spiller skal gradere sin egenopplevde oppfattelse av utmattelse fra 0 (hvile) til 10 (maksimal anstrengelse) (Foster et al., 2001). Etter en trening eller kamp multipliseres spillernes sRPE med varigheten på økten for å få spillerens opplevde (interne) belastning kvantifisert. sRPE har blitt kjent som en god indikator på spillernes interne belastning, mye på grunn av den tar høyde for faktorer som skader, sykdom, kampprogram og dag til dag variasjon til spillernes fysiske form (Impellizzeri, Rampinini, Coutts, Sassi, & Marcora, 2004). Det er metodiske begrensninger ved bruk av sRPE, knyttet til at det er subjektivt, men også begrensninger i å regulere forholdet mellom intern (intensitet) og ekstern (varighet) belastning i ulike økter (Soligard et al., 2016). Og studier har vist at sRPE-metoden ikke egner seg til å oppdage små endringer i treningsintensiteten under korte sporadiske økter (Scott, Black, Quinn, & Coutts, 2013). Til tross for denne begrensningen, forblir sRPE en gyldig metode for å kvantifisere treningsbelastning i ulike idretter (Haddad, Stylianides, Djaoui, Dellal, & Chamari, 2017).

Den tredje dimensjonen på hvordan en måler treningsbelastning er den absolutte/relative treningsbelastningen. Absolutt belastning er summen av belastning påført utøveren over en gitt periode. Relativ belastning tar høyde for hvor fort belastningen øker/blir påført utøveren, og utøverens tidligere belastning. Det blir ofte uttrykt som ratio av belastning mellom to ulike perioder, eller forholdet mellom en ny tidsperiode (vanligvis de foregående 7 dagene) og den forgående tidsperioden (vanligvis siste 4 uker/28 dager). Dette blir kalt for Acute Chronic Workload Ratio (ACWR) og ble presentert i 2016 som en modell som vurderer sammenhengen mellom endringer i treningsmengde og risikoen for utvikling av skade (Hulin et al., 2014).

Selv om en høy treningsbelastning er assosiert med en høyre skaderatio (Gabbett & Domrow, 2007; Gabbett & Jenkins, 2011; Harrison & Johnston, 2017), er resultatene tvetydige. Forskning viser også at høy treningsbelastning kan ha en skadeforebyggende effekt (Gabbett, 2016; Hulin et al., 2014). Generelt antydes det at sannsynligheten for skade er lav når ACWR forholdet er innenfor området på 0,85 til 1,35, og risikoen blir mer enn det dobbelt når forholdet overstiger 1,5 (Soligard et al., 2016). Men en slik anstrengelse har også blitt kritisert i en rekke studier (Hulin, 2017; Lolli et al., 2019; Sampson, Fullagar, & Murray, 2017).

Gitt at høy treningsbelastning kan oppnås på ulike måter (volum, intensitet, og frekvens på treningene), vil det være usannsynlig at all «høy treningsbelastning» har identisk skaderisiko (Gabbett, 2016). Høy treningsbelastning i seg selv er ikke en sikker risikofaktor for økt skaderisiko. Større mengde med høyintensitets akselerasjoner, og spill spesifikke aktiviteter kan være viktig med tanke på å trene de fysiske kvalitetene som trengs for å prestere på et høyt nivå. Det kan også være en mulig beskyttende faktor for spillerne med tanke på å utvikle en skade (Gabbett, 2016). Mens en signifikant endring i belastning som fører til akutt høy ukentlig belastning og store uke-til-uke endringer i belastningen har blitt assosiert med økt skaderisiko (Malone et al., 2017b; Matthew J. Cross, 2016; Soligard et al., 2016). Det vil si at det ikke er den høye treningsbelastningen som er problemet, men hvordan høy treningsbelastning blir implementert.

Selv om forståelsen av sammenhengen mellom treningsbelastning og risikoen for skade i idrett øker, er det fortsatt utfordringer og begrensninger til denne forskningen. I fotball er det mangel på studier, og det er få studier gjennomført på mannlige profesjonelle fotballspillere (Malone et al., 2017b; McCall, Dupont, & Ekstrand, 2018). Med få

studier, og studier som ikke har data over flere sesonger får en ikke registrert mange nok skader. For å kunne finne en moderat til sterk assosiasjon i studier som ser på risikofaktorer blir det anbefalt å ha 20 til 50 skader (Bahr & Holme, 2003).

2.4.1 Validitet og reliabilitet av GPS-enheter

En studie gjennomført av Rampinini et al. (2015) på 10 Hz GPS rapportere om gode resultater på variasjonskoeffisient (VK) for totaldistanse (VK=1,9%) og høyhastighetsløping (>15 km/t; VK=4,7%). Studien viste også en svakere VK (VK=10,5%) for høyhastighetsløp >20 km/t, noe som betyr at validiteten reduseres når hastigheten øker. Johnston, Watsford, Kelly, Pine, og Spurrs (2014) fant valide resultater når de testet enhetene på en simulator som skulle simulere ballspillbevegelser, resultatet var <1% feilmargin på totaldistansen. Castellano, Casamichana, Calleja-González, Román, og Ostojic (2011) rapporterte at distansedata var nøyaktig, men at GPS-enhetene underestimerte den virkelige distansen. GPS-enhetene viste seg å ha høy grad av intra- og interreliabilitet, og at nøyaktigheten økte ved lengre distanse. Johnston et al. (2014) fant at det var en økende feilmargin når hastigheten økte. For høyhastighetsløp (>20 km/t) var feilmarginen over 10% på 30-metersløp. Rampinini et al. (2015) testet høyhastighetsløp i en løype på 70 meter (2x35 meter med vending). Resultatene viste en KV på 4,7%. For akselerometerdata viste Gallo, Cormack, Gabbett, og Lorenzen (2015) og Polglaze, Dawson, Hiscock, og Peeling (2015) en høy korrelasjonskoeffisient på henholdsvis ($r=0,97$) og ($r=0,86$) mellom total distanse og akkumulert PlayerLoad™ i idrettene Australsk fotball og landhockey. Scott, Scott, og Kelly (2016) fant at alle GPS-enheter, uavhengig av innsamlingsfrekvens, er i stand til å monitorere utøvere, men at det per dags dato ser ut til at 10Hz GPS-enheter er de mest valide og reliable.

2.5 Maskinlæring

Idrett utvikler seg fra et felt som i hovedsak har vært avhengig av menneskelig kunnskap, til et område drevet av dataanalyse (Cintia, Giannotti, Pappalardo, Pedreschi, & Malvaldi, 2015). Idrettsanalyse er et nytt felt, som modellerer idrettsprestasjonen med en vitenskapelig tilnærming ved bruk av ulike metoder og teknikker som maskinlæring, kinesiologi, statistikk og teorier (Miller, 2015). Analyse av idrett har endret hvordan ulike idrettsgrener utføres. Før var det vanlig å stole på intuisjon og magefølelsen for å

ta avgjørelser rundt laget. Ved hjelp av maskinlæring og statistiske teknikker kan trenere finne fordelaktige tilnærminger og ta viktige avgjørelser (Robert P. Schumaker., 2010).

Analyse av data kan spille en viktig rolle for å forebygge skade i fotball. Ved å bruke data til å forutse om en spiller kommer til å bli skadet før skaden skjer, kan en gjøre nødvendige endringer eller individuelle tilpasninger for å redusere risiko for skade. En slik tilnærming vil være til fordel for aktive utøvere (Stern, Hegedus, & Lai, 2020). Analyse av spillernes data er avgjørende for å sikre at spilleren har en tilpasset mengde treningsbelastning og passende restitusjonstid mellom kamper og treningsøkter (Colby, Dawson, Heasman, Rogalski, & Gabbett, 2014).

Teknologien som brukes i det mest innovative miljøet i helsevesenet kalles kunstig intelligens (KI), og et annet navn som ofte blir brukt i sammenheng med dette er maskinlæring (ML). ML beskriver bruken av en algoritme for å identifisere mønster i store datasett. Klinikere vil i fremtiden måtte jobbe med et informasjonsnettverk som er i en skala langt utenfor mennesker kapasitetsevne, og dermed vil det bli nødvendig å bruke intelligente maskiner for å analysere og tolke de komplekse samhandlingene en finner i et datasett (Frankish Keith., 2017; Michael, uå).

Moderne tilnærminger ved bruk av KI har vist seg å ha et høyere suksessnivå når de modellerer usikre utfall. Mange av disse fremskrittene er knyttet til fremgang i underdomene ML, som beskriver en algoritmes evne til å lære. Svarene produsert av ML-algoritmer er beslutninger laget av statistiske analyser av store datasett. Svarene er uttrykt som sannsynligheten for et forhold mellom variabler. Dette vil si at svaret man får ikke nødvendigvis er det korrekte svaret, men mer som en antagelse på hva svaret kan være (Frankish Keith., 2017; Michael, uå).

I en 3-steps prosess med å velge datasett, lage en algoritme og evaluere og finjustere modellen i etterkant, så er ingenting mer kritisk enn kvaliteten i dataen man benytter i modellen (Verghese, Shah, & Harrington, 2018). Det oppstår ulike utfordringer når det kommer til selve dataen, og alle disse utfordringene kan påvirke valgene knyttet til implementeringen av ML.

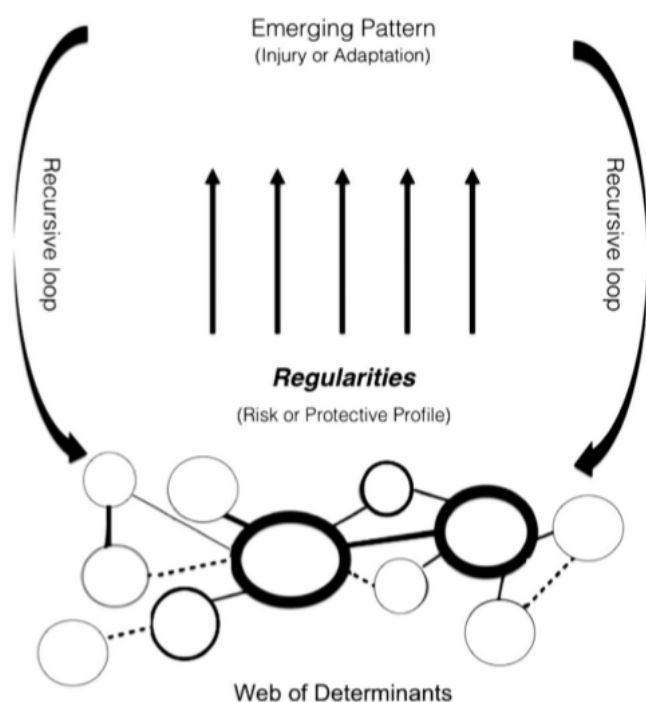
Ubalanse i dataen kan oppstå når treningsdatasettet inneholder flere eksempler på noen kategorier enn andre. For eksempel kan det være lurt å ha en algoritme som kan forutse sannsynligheten for en sjelden skade i et fotballag. Du kan trene algoritmen i et datasett

som inkluderer nesten ingen typer av denne skaden (fordi den er veldig sjelden), men du ender opp med å nesten oppnå en perfekt score på resultatet ditt. Dette skjer fordi algoritmen forutser null forekomster av skaden. I slike tilfeller vil algoritmen ha en tilsynelatende god prediktiv evne, selv om den er nøyaktig, vil nytteverdien være begrensende. Eksempler som dette viser at man vil ha behov for trenere og klinikere som skjønner hvordan ML fungerer, og på den måten tolke resultatene fra analysen.

Å predikere skade er en av de mest utfordrende problemene innen idrett, samtidig er det en nøkkelfaktor for å forebygge skader. Undersøkelser som har sett på årsaken til skader i idrett har tradisjonelt sett hatt et reduksjonistisk syn på hvordan man ser på skaden. Skaden har ofte blitt forenklet til enheter og da analysert skadens grunnleggende deler. Årsakssammenhengen er blitt sett på en lineær måte (Bittencourt et al., 2016). En reduksjonistisk tilnærming er avhengig av korrelasjons- og regresjonsanalyser, og til tross for mange studier som har prøvd å predikere idrettsskader, har denne type forskning vært begrenset med sin evne til å identifisere prediktive faktorer. Majoriteten av helseproblemene er komplekse. Disse faktorene kan være koblet sammen på en ikke-lineær måte, som kan bety at store endringer i en variabel ikke nødvendigvis får en stor effekt på utfallet. Og en liten endring hos noen determinanter kan føre til store, og i noen tilfeller uventet konsekvenser på utfallet (Higgins, 2002). På bakgrunn av dette så trengs det en bredere tilnærming, for å kunne forstå den komplekse sammenhengen mellom risikofaktorer, prediktive faktorer og skade (Mendiguchia, Alentorn-Geli, & Brughelli, 2012; Quatman, Quatman, & Hewett, 2009).

Bittencourt et al. (2016) lagde en alternativ skadeetiologi-modell der faktorene blir satt i et komplekst system. Det som blir sett på som viktige faktorer i et komplekst system, resulterer i en gruppe av determinanter. Faktorene interagerer med hverandre på uforutsigbare og uplanlagte måter (Rickles, Hawe, & Shiell, 2007). Disse komplekse interaksjonene oppstår som følge av systemets egen historie (eks. utøverens liv). Interaksjonene mellom determinantene danner regelmessigheter, som enten kan være en risiko eller beskyttende faktor for utøveren, som tilslutt fører til skade eller adaptasjon hos utøveren. Kompleksitet og prediksjon krever at en aksepterer et visst nivå av usikkerhet, med hensyn til at skaden som eventuelt oppstår, ikke kan garanteres (Doll & Trueit, 2010). Det kan sammenlignes med et værvarsel, prediksjonen forbedres over tid, men feil er alltid en mulighet. For å imøtekomme dette usikkerhetsproblemet med KI forslår Bittencourt et al. (2016) at idrettsskadeforskningen må konsentrere seg om å

analysere de observerbare regelmessighetene som oppstår fra de eksiterende interaksjonene mellom faktorer, og ikke de allerede etablerte risikofaktorene i seg selv.



Figur 6: Kompleks modell for idrettsskader. Gruppen med sirkler nederst skal inneholde ulike variabler, disse utgjør determinantene. Variablene med de mørke sirkler har flere interaksjoner enn variablene med lysere sirkler og har en større påvirkning på utfallet. De med stiplet linje representerer svakere interaksjoner, mens de med mørk linje indikerer sterkere interaksjon mellom variablene (Bittencourt et al., 2016).

Ved hjelp av bærbar teknologi som hjertefrekvensmåler og GPS-enheter, er det mulig med sanntid og kontinuerlig data under treninger og kamp. Ved å bruke nye målinger kan man konstruere nye dynamiske modeller for å predikere skader. For å få en mer fullstendig oversikt over determinanter så anbefales det at en i tillegg til GPS-data samler inn data på selvrapporing om stress, angst og mestringsevne, søvnkvalitet og ernæring, samt at en gjennomfører mobilitetstesting og testing av stabilitet, kraft og motorisk kontroll (Stern et al., 2020).

Bruken av kunstig intelligens åpner opp et interessant perspektiv for å forutsi skaderisiko og ulike prestasjoner i lagidretter (Claudino et al., 2019). I studien til Claudino et al. (2019) lagde de en oversikt over de mest brukte metodene i bruken av kunstig intelligens når det kom til å vurdere skaderisiko og predikere prestasjoner hos

utøvere i lagidretter. Metodene med størst potensial var Artificial neural networks, Decision tree classifier, Markov process og Support vector machine.

Maskinlæring har blitt brukt til å predikere utfallsvariabler i en mengde ulike områder. Studien til Lampham (1995) var en av de første studiene som brukte kunstig intelligens i sine analyser. I studien viste de kunstig intelligens fordeler i ulike settinger. Bruken av KI i idrettsrelatert forskning har blitt veldig populært de siste årene (Farmer et al., 2012), maskinlæring har blitt brukt til å predikere resultater på kamper i baseball, basketball og fotball (Joseph, Fenton, & Neil, 2006; Soto-Valero, 2016). Et økende antall studier har i nyere tid prøvd å predikere skader ved å bruke allerede etablerte risikofaktorer (Carey et al., 2018; Ruddy et al., 2019).

Maskinlæring innebærer å lage algoritmer som kan lære av data og komme med prediksjoner. Algoritmen skal kunne finne forutsigelser uten å bli programmert hva den skal se etter eller hvor den skal se etter dem (Ruddy et al., 2019). Navnet maskinlæring blir brukt fordi metoden lærer fra en prosess som kalles «trening». De to vanligste metodene innen ML er supervised learning og unsupervised learning (Liu, Chen, Krause, & Peng, 2019). Unsupervised learning er en prosess hvor prediksjonene fra algoritmen er laget på datasettet uten noen korresponderende utfallsvariabler. I prospektive studier som ser på skader, er utfallsvariabelen (skade/ ikke skade) som oftest kjent. Derfor er supervised learning en mer relevant metode å bruke på forskning tilknyttet skader. Supervised learning er en prosess hvor et datasett med en kjent utfallsvariable blir brukt som treningsdata. Algoritmen øver seg på å identifisere mulige sammenhenger og til å predikere de samme variablene som den skal finne i testdataen. Treningsdataen blir brukt til å bygge modellen, mens testdataen blir brukt til å måle modellens evne til å predikere utfallet. Det er viktig at testdataen ikke er den samme dataen som ble brukt til å bygge modellen. Da kan det oppstå et problem som kalles «overfitting». Dette skjer når treningsdataen som er valgt passer for bra med testdataen. Dette reduserer generaliserbarheten til modellen, og reduserer modellens evne til å predikere på usett testdata (Kotsiantis, 2007).

En større mengde med treningsdata vil gjøre det mulig for algoritmen å lære og identifiserte komplekse og ikke-lineære sammenhenger, hvis dette finns i datasettet (Ruddy et al., 2019). Datasettet som algoritmen skal bygges opp på kan deles inn på flere måter. Et eksempel kan være at datasettet består av data fra flere sesonger, da kan de tidligere sesongene brukes som treningsdata til å predikere utfallene til den

kommende sesongen. Det er ingen konsensus på hvor mye data som bør brukes som treningsdata, men en studie har foreslått at man brukes mellom 40 til 80 prosent av all innsamlet data som treningsdata. Dette vil mest sannsynlig resultere i en optimal prediktiv evne hos algoritmen (Dobbin & Simon, 2011). Det finns flere forskjellige algoritmer som kan brukes når man bygger en prediktiv modell (Kotsiantis, 2007). Hver algoritme har forskjellige underliggende matematiske funksjoner, samt et unikt sett med parameter som kan kontrolleres for å bestemme hvordan algoritmen samhandler og lærer av datasettet. Noen algoritmer er mer robuste og har mindre sjanse for å bli påvirket av små nyanser i datasettet, mens andre er mer komplekse og sensitive. Forskjellige typer algoritmer passer til forskjellige type data, men det er ofte praktisert at en implementerer forskjellige algoritmer og sammenligner resultatene. I denne studien ble de vanligste algoritmene for binær prediksjon brukt, logistisk regresjon, decision tree, random forest og support vector machine.

En stor mengde med treningsdata forbedrer modellens evne til å identifisere mønster og gjøre meningsfulle prediksjoner. Det er vanskelig å estimere mengde data som trengs for å oppnå en mer presis modell. En studie av Peduzzi, Concato, Kemper, Holford, og Feinstein (1996) mente at det bør være opptil 10 skader per variabel for at prediksjonen skal kunne gi mening. Dessverre så er evnen til å samle inn store datasett på idrettsskader en begrensning innen idrettsmedisinsk forskning (van Dyk, van der Made, Timmins, Opar, & Tol, 2018). Gitt begrensingene ved små datasett (få skader), er det fortsatt sannsynlig at komplekse tilnærminger gir mer informative resultater enn den reduksjonistiske tilnærmingen (Bittencourt et al., 2016).

2.6 Belastningsstyringscyklusen

Treningsbelastning blir monitorert med mål om å ta evidensbaserte beslutninger om treningsbelastningen man påfører spilleren, dette er viktig for å optimalisere prestasjonsevnen til spillerne og samtidig redusere skaderisikoen (Akenhead & Nassis, 2016; Weston, 2018).

Forskningen innen fotballvitenskap har økt gjennom de siste årene, noe som har resultert i utviklingen av ny teknologi for overvåking av treningsbelastningen (Akenhead & Nassis, 2016). Ved siden av teknologisk utvikling har mengden forskning

som bruker GPS-enheter økt i løpet av det siste tiåret (Malone, Lovell, Varley, & Coutts, 2017a; Rico-González, Los Arcos, Nakamura, Moura, & Pino Ortega, 2019).

Utviklingen innen teknologi og ulike analysemetoder har ført til nye muligheter i fotballen, nå kan man overvåke treningsbelastningen til spilleren ved hjelp av GPS-enheter og annen mikroteknologi. GPS-enhetene produserer en mengde data som gjør at man kan analysere treningsbelastninger mer detaljert. Det er foreløpig ingen enighet om hvilke variabler som er mest nyttige, eller hvordan en skal analysere longitudinelle data fra spillergruppen (Akenhead & Nassis, 2016).

I studien til Akenhead og Nassis (2016) svarte 37 av 41 lag at de brukte Microsoft Excel til å analysere og lage rapporter av dataen, og 23 av lagene brukte også et ytterligere program (eks. openfield). Lagene som svarte på spørreundersøkelsen til Akenhead og Nassis (2016) sa de brukte de ulike variablene til treningsbelastning til å planlegge spesifikke tidsperioder gjennom sesongen. Flere av lagene brukte dataen til planlegging av akutt/kronisk treningsbelastning, med perioder fra 3 dager til 6 uker. Studien til Dalen-Lorentsen (2019) viser at forholdet mellom akutt/kronisk treningsbelastning og helseproblemer er avhengig av metoden som blir brukt. Studien viser til at det ikke er noe konsensus på hvordan forholdet mellom akutt/kronisk treningsbelastning og helseproblemer skal analyseres og kalkuleres.

I et fotballag er det ofte mange forskjeller når det kommer til ikke-modifiserbare risikofaktorer som alder, skadehistorikk og treningshistorikk. Spilleren har ofte forskjellig treningsbelastning og kampbelastning. Dette påvirkes av at spillerne har ulike posisjoner på banen, ulik spilletid og hvem en spiller mot (Akenhead & Nassis, 2016). Dette gjør at både spillerne og trenerne er interessert i hver enkelt spiller sin belastning, endring i belastning og skaderisiko til enhver tid. Idrettsforskere har kritisert tester og tiltak som er basert lineære modeller og multiple regresjonsmodeller (Bahr, 2016; Freckleton & Pizzari, 2013). Kritikerne har argumentert med at ingen test eller måling har vist sterk nok følsomhet og spesifisitet til å kunne forutsi skade, og at de stiller spørsmålsteget om skadescreening og prediksjon av skade er for komplisert for de lineære modellene som har blitt brukt tidligere (van Dyk & Clarsen, 2017).

Hovedfunnene til Akenhead og Nassis (2016) viser at det ikke er noen universell tilnærming til hvordan en skal bruke og analysere treningsbelastningen til spilleren. De

konkluderer med at den opplevde effekten av skadeforebygging og forbedring av prestasjon er lavere enn forventet.

Ved å se på spillernes motstandsdyktighet til skader som et dynamisk system, kan vi klare å forbedre prediksjonen av skader. Det kan føre til bedre metoder til å samle inn data, og en bedre forståelse på forholdet mellom de ulike variablene som har en assosiasjon til skade (Stern et al., 2020). Det er med bakgrunn i dette vi ønsket å lage en modell som bruker kunstig intelligens til å finne mønster i den store datamengden man får ut fra GPS-enhetene. På den måten bruker man modellen til å analysere dataen, som legger grunnlaget for hvordan en skal styre belastningen til spillerne gjennom en sesong.

3. Metode

Datainnsamlingen og metoden til denne masteroppgaven ble gjort med en kombinasjon av datainnsamling og praksis i en fotballklubb. Maskinlæringsmodellen som er brukt i denne masteroppgaven ble utviklet i samarbeid med UIO, og var grunnlaget for en masteroppgave i informatikk.

3.1 Studiedesign

Studien var en longitudinal kohortstudie av et mannlig norsk eliteserielag i fotball der vi registrerte forskjellige variabler for treningsbelastning og eventuell skade (r) på enkeltspillere gjennom en fotballsesong. Datainnsamlingen startet 31. mars 2019 og ble avsluttet ved sesongslutt 1. desember 2019. Oppkjøringen for sesongen startet i januar samme år, og jeg fikk tilgang på GPS-data fra perioden jeg selv ikke var til stede. I denne perioden ble det også samlet inn treningsbelastningsdata og skaderegistreringer fra alle treninger og kamper som jeg har benyttet i denne oppgaven.

3.2 Utvalg

Deltakerne i studien ($n=32$, alder: $25 \pm 4,6$) ble rekruttert fra førstelagsstallen til en norsk eliteserieklubb. Vi hadde opprinnelig data på 38 spillere, men data fra 6 spillere ble ekskludert grunnet for få registrerte treninger (< 7). Rekrutteringen av spillere til deltagelse i studien skjedde i samarbeid med klubben. Spillerne er alle voksne, men med varierende erfaring fra eliteseriespill. Spillerne fikk muntlig og skriftlig informasjon om studien før oppstart, og signerte på et samtykkeskjema hvis de ønsket å delta. Det ble totalt registrert data fra 2848 treninger, i gjennomsnitt 89 økter (± 45) per spiller. Vi registrerte totalt 34 IK-skader (tabell 3) gjennom sesongen. Kontaktskader er ikke inkludert i det statistiske grunnlaget.

3.3 Datainnsamling

Datainnsamling foregikk på Marienlyst Stadion i samarbeid med fysisk trener i klubben, under trening på treningsdagene KD-4, KD-3 og KD-2, og i tillegg under kampen til den tilhørende uken. Under bortekamper og resterende treningsdager, hvor jeg ikke var

til stede var det fysisk trener som samlet inn dataen. Det var det fysiske-trenerapparatet sammen med meg som hadde ansvar for at alle spillerne tok på seg vesten og den tilhørende GPS-enheten før treningen startet. Sammen hadde vi også ansvar for å skru på enhetene før trening, og samle de inn GPS-enhetene etter at treningsøktene var avsluttet. Øktene ble registrert ved hjelp av en datamaskin som håndterte data fra Catapult-systemet. Datamaskinen var tilkoblet en direktesendt antenne (TRX; Catapult sports, Australia; Firmware version 7.22). Etter treningsøktene ble alle GPS-enhetene plassert i en docking-stasjon (S5 Charge case; Catapult Sports, Queensland Australia). Data ble deretter overført til datamaskinen og videre til Catapult nettskyen.

3.4 Skadedefinisjon

I vår studie registrerte vi to ulike typer ikke-kontaktskader: *Ikke-kontakt skade (IK)* og *ikke-kontakt skade med fravær (IK-F)*.

IK ble definert som en skade uavhengig symptomer som ikke førte til fravær fra trening eller kamp. IK-F ble definert som en skade som resulterte i fravær fra neste trening eller kamp. En spiller var registrert som skadet helt til han ble klarert spilleklar av det medisinske støtteapparatet i klubben.

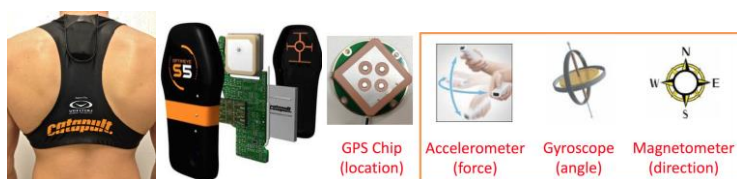
3.4.1 Skadekategorier

Vi deler skadene inn i to kategorier: *Akutte belastningsskader* og *belastningsskader*. Akutte belastningsskader skyldes at muskler, sener eller ledd har blitt overbelastet i en gitt situasjon eller øyeblikk. (Bahr, McCrory, Laprade, Meeuwisse, & Engebretsen, 2014).

Belastningsskade oppstår etter at en muskel eller kroppsdel har blitt ensidig belastning over tid. Typisk eksempel er akilles-tendinopathi. I slike tilfeller har en ikke et bestemt skadetidspunkt (Bahr et al., 2014; Sverre Mæhlum, 2017). Men en belastningsskade kan også oppstå helt uventet uten forvarsel, dette kan fort feiltolkes som en akutt skade, men at det egentlig er et resultat av lengre tids overbelastning (Windt & Gabbett, 2017). Et eksempel på dette er en hamstringstrekk (Sverre Mæhlum, 2017).

3.5 Utstyr og målemetode

Datainnsamlingen ble gjennomført ved bruk av en GPS-enhet med innebygd 1000Hz akselerometer, gyroskop og magnetometer (Catapult OptimEye X4). GPS-enheten som brukes i dette prosjektet, er en oppgradert utgave av MiniMaxX S4 (Catapult, Sports, Australia). GPS-enheten ble plassert i en spesialdesignet vest, hvor enheten ligger i en lomme mellom skulderbladene (figur 7) (Catapult Sports, Australia). Tallene fra GPS-brikkene ble bearbeidet og hentet ut fra programmet Catapult Openfield.



Figur 7: Eksempel på en GPS-enhet med innebygd akselerometer og på en spesialdesignet vest som tilsvarer den spillerne brukte.

3.6 Dataeksport

Data ble overført fra GPS-enhetene til datamaskinen, dataen ble analysert for å avdekke om noen av forsøkspersonene manglet data eller hadde ufullstendig dataregistreringer. Disse målingene ble senere ekskludert fra analysen. Dataen ble lastet inn i programmet Openfield (version 1.17.0, Catapult Sports, 2018, BUILD#30874). Når sesongen var avsluttet, ble alle treninger og kamper lastet ned i ett-sekunders intervaller, øktene ble lagret som CSV-filer og sendt til samarbeidspartneren vår som lagde algoritmen og satte opp modellen vi brukte til å gjennomføre prediksjonsanalysene. Januar-dataene ble ekskludert, fordi det fantes bare en økt i januar med GPS-data på, og den ble gjennomført lenge før den første økten i februar. I tabell 1 er det en oversikt over variablene med beskrivelser av hvilke variabler som ble hentet fra GPS-dataen.

Tabell 1: Oversikt over hvilke variabler som ble hentet fra GPS dataene, hva de måler og hvilken måleenhet som benyttes.

Navn	Forkortelse	Hva den måler	Måleenhet
PlayerLoad	PL	Den totale summen av akselerasjonene på tvers av alle akser	Arbitrary unit (AU)

Acceleration Load	AC	Den totale summen av akselerasjonsverdiene for den angitte perioden. (Totale mengden med hastighetsendringer)	Arbitrary unit (AU)
Total distanse	TD	Totale distansen spilleren har beveget seg (Uavhengig av hastighet)	Meter (m)
V4 distanse	V4	Den totale distansen utøveren har løpt med hastighet mellom 20-24,99 km/t	Meter (m)
V5 distanse	V5	Den totale distansen spilleren har løpt med hastighet over 25 km/t	Meter (m)
Høyhastighetsløp	HHL	Den totale distansen spilleren har løpt med hastighet over 20 km/t	Meter (m)

3.7 *Catapult OptimEye X4*

Catapult OptimEye X4, er en elektrisk GPS-enhet som veier 66,7 gram. Den er 96 mm høy, 55mm i bred og 14 mm tykk (Catapult Sports, Queensland Australia). Enheten samler på 2,4 GHz og sender data live til en datamaskin. På denne måten kan den brukes til måling av bevegelser i sanntid (Catapult Sports, 2018b). Malone et al. (2017a) argumenterte i sin studie at GPS-enhetene må ha 4 satellitter til enhver tid tilkoblet for at det skal gi et adekvat signal, men det ideelle er 6 satellitter. Data som hadde mindre en 4 satellitter tilkoblet ble ekskludert fra analysen vår.

3.8 *Player Load™*

Den mest kjente definisjonen på ytre belastning ved siden av total distanse er Catapult Sports Playerload™. Playerload™ er et akselerometerbasert måleinstrument som måler den eksterne belastningen hos utøvere i lagidretter (Luteberget & Spencer, 2017). Playerload™ summerer kreftene i alle tre akser (X, Y, Z) som påføres utøveren under

fotballtrening og kamp. Gjennom en likning gir den en indikasjon på hvor belastende treningen eller kampen har vært.

Utregningen av playerloadTM

$$Playerload^{TM} = \frac{(a_{y1} - a_{y-1}) + (a_{x1} - a_{x-1}) + (a_{z1} - a_{z-1})}{100}$$

a_y = fremover akselerasjon

a_x = sideveis akselerasjon

a_z = vertikal akselerasjon

3.9 Datasett

Dataen vi benyttet er hentet fra 3005 innsamlede økter fra 38 spillere. Seks av spillerne ble ekskludert på grunn av for få registrerte treninger. I rådataene fant vi 174 økter med urealistiske tall, og data fra disse øktene ble også ekskludert. I ett tilfelle registrerte vi en spiller som løp 50 km på ett sekund, som åpenbart var en feilmåling. Tallene fra disse øktene kan forklares med feil på GPS-brikkene eller at de har skrudd seg av grunnet strøm mangel under trening eller kamp. Problemene med GPS-dataen ble løst på to ulike måte. Den ene metoden var å fjerne økter med urealistiske tall, den andre metoden var å lage gjennomsnitt til en spiller, basert på resten av spillergruppen.

3.10 Maskinlæringsmodellene

Fire ulike typer maskinlæring ble brukt til å prøve og predikere skade. Hver modell ble testet 1000 ganger, og ble presenter som et gjennomsnitt av alle 1000 testene. Hver av disse modellene bruker fire algoritmer som varierer i kompleksitet og hvordan disse tolkes. I det følgende avsnitt beskriver jeg modellene og teknikkene som er brukt. For en 100 prosent korrekt forklaring av metoden som er brukt for å lage denne modellen, anbefaler jeg å lese masteroppgaven *The use of Data Mining for Predicting Injuries in Professional Football Players* av Garth Theron (Theron, 2020).

Support vector machine (SVM): SVM er basert på en algoritme som skal finne et «hyperplan» som best mulig deler datasettet inn i to klasser. Desto lengre datapunktene er fra hyperplanet desto sikrere er vi på at datapunktet har blitt kategorisert riktig.

Logistic regression (LR): Denne modellen egner seg godt til den binære delen av problemet, som modelleres i denne oppgaven. Det er et vanlig valg for modellering av utfallet for skade. Ved å bruke logistisk funksjon produserer denne algoritmen en S-formet prediksjonskurve som bregner prediksjon til en verdi mellom 0 og 1 (Gareth James., 2014).

Decision tree (DT): Denne klassifiseringen er en ikke-parametrisk tilnærming. Ved bruk av DT er modellen programmert til å lære seg enkle beslutninger (skade/ikke-skade) som påvirkes av de ulike prediktorvariablene. DT er ansett som en ideell fremgangsmåte for få innsikt i sammenhengen mellom skade og treningsbelastning på grunn av den relativt enkle måten den kan tolkes på.

Random forrest (RF): Det er en metode som bygger videre på DT, i et forsøk på å få til en enda sterkere/bedre prediksjonsmodell. Algoritmen fungerer på en måte der man prøver å bygge flere DT.

3.11 Valg av parameter

Som nevnt innledningsvis så var oppsettet av maskinlæringsmodellene basert på metoden til Alessio Rossi (2018). I studien til Rossi identifiserte de tre parameter som var relevante for skadeprediksjon; estimert rullende gjennomsnitt for spillerens høyhastighetsløp (HSR^{EWMA}), gjennomsnittlig standardavvik ratio hos en spillers totale distanse ($Dist^{MSWR}$) og estimert vektet rullende gjennomsnitt for tidligere skade (PI^{EWMA}). I denne oppgaven bruker vi en kombinasjon av ulike teknikker og funksjoner fra studien til Alessio Rossi (2018) for å teste reproduserbarheten av disse parameterne på et nytt datasett.

3.12 Parameter

3.12.1 Mean Standard deviation Workload Ratio (MSWR)

MSWR er ofte brukt for å kvantifisere hvor monoton arbeidsbelastningen til en spiller er (Carey et al., 2018). MSWR er definert som ratioen mellom gjennomsnittet og standardavviket hos hver enkelt spillers arbeidsbelastning h som er påført spilleren i en spesifikk periode. Med arbeidsbelastning mener vi ulike parameter som høyintensitetsløp, PlayerLoad, akselerasjonsbelastnings data, V4 distanse og V5 distanse. MSWR-verdier representerer variasjonene hos en spillers arbeidsbelastning.

Høye MSWR-verdier indikerer en mer monoton arbeidsbelastning i den spesifikke perioden.

$$MSRW_t = \frac{\mu t}{\sigma t}$$

Hvor:

μt er gjennomsnitt av en parameter (eks total distanse) i en gitt periode t

σt er standardavvik hos en parameter for en gitt tidsperiode t

3.12.2 Exponential Weighted Moving Average (EWMA)

For å se på den seneste treningsbelastningen til en spiller, beregner vi eksponentielle vektete rullende gjennomsnitt (EWMA) av de seneste treningsøktene. Dette ble utført gjennom at EWMA vektet viktigheten av ny belastning, hvor nyere treningsdata/belastningsdata er vektet mer enn eldre treningsdata/belastningsdata. Desto nyere en verdi er, desto mer blir den vektet i utregningen. Det rullende gjennomsnittet blir beregnet på følgende måte:

$$EWMA_{idag} = Belastning_{idag} \times \lambda + ((1 - \lambda) \times EWMA_{igår})$$

Hvor $0 \leq \lambda \leq 1$

λ representerer graden av forfall, hvor eldre verdier reduseres fortere enn nye verdier.

Representert som: $\lambda = 2 / (N + 1)$

Hvor N er referert til omfanget, og representerer den valgte forfallskonstanten.

3.12.3 Skadeparameter

Skadeparameteret i denne oppgaven beregner antall tidligere registrerte skader hos en spiller. I tillegg til hvor mange dager det er siden spilleren kom tilbake i spill siden forrige skade. I utregningen modifiserer vi EWMA og kan beskrives som et rullende gjennomsnitt fra tidligere og nåværende observasjoner.

$$Z_i = \lambda + \sum_{j=0}^{i-1} (1 - \lambda)^j X_{i-j} + (1 - \lambda) i Z_0$$

Hvor:

Z_i er den kalkulerede EWMA for økter i

X_i er den registrerte verdien for økt nummer i

λ er en verdi fra 0 til 1

Z_0 er EWMA startverdi

Er skadeparameterverdien null, betyr det at spilleren aldri har vært skadet. En verdi over null indikerer at en spiller har hatt minst en skade i løpet av sesongen. For spillere som kommer tilbake fra en skade så øker verdien på skadeparameteret for hver dag. For spillere som har vært skadet en gang så øker verdien fra 0 helt til den eventuelt når 1. Samme for spillere som har blitt skadet for andre gang, da øker verdien på featuren fra 1 og opp til 2. Hvor fort featuren øker i verdi er forårsaket av λ , hvor høyere verdi gir uttrykk for en raskere økning i skadeparameterverdien.

3.12.4 Presisjon

Presisjon beskriver brøkdelen av antallet klassifisert som positive og som faktisk er positive (Menardi & Torelli, 2014). Den fungerer som en indikator på modellens pålitelighet. Ved prediksjon av skade vil en modell med høy presisjon identifisere skader med stor sannsynlighet. Ved en lav presisjon vil modellen predikere mange falske alarmer. Presisjon blir referert til som «sann positiv». Det måler den andelen av spillere som korrekt ble klassifisert som skadet (Ruddy et al., 2019).

$$Presisjon = \frac{Sann\ positiv}{Sann\ positiv + falsk\ negativ}$$

3.12.5 Sensitivitet

Det er en indikasjon på modellens evne til å identifisere type skade. Jo høyre sensitivitet, jo bedre er modellen til å oppdage type skade. En lav sensitivitet vil

resultere i at modellen har en manglende evne til å predikere om en spiller kommer til å bli skadet.

$$\text{Sensitivitet} = \frac{\text{Sann positiv}}{\text{Sann positiv} + \text{falsk negativ}}$$

3.12.6 F1 score

F1-score er et gjennomsnitt mellom sensitivitet og presisjon der en kombinerer de to til ett mål. Kan beskrives som modellens nøyaktighet.

$$F1 - score = 2 * \frac{\text{Sann positiv}}{\text{Sann positiv} + \text{falsk negativ}}$$

3.12.7 Areal under kurven (AUK)

Receiver Operating Characteristics (ROC) er et verktøy som ofte blir brukt for å evaluere hvor godt en modell klarer å predikere ubalanserte data. Det er et grafisk plott hvor en plotter modellens «true/positive rate» (sensitivitet) mot modellens «false/positive rate» (Se figur 7). AUK gir et aggregert mål på prestasjonen til modellen og måler det todimensjonale arealet under ROC-kurven. Verdi nærmere 1 er assosiert med en god prediksjon, mens verdier rundt 0,5 tilsvarer gjetting (Menardi & Torelli, 2014). For å sette det i perspektiv så vil en AUK-score på 0.5 tilsvare at en flipper en mynt (Bahr, 2016).

3.12.8 Adaptere syntetisk data

Adaptere syntetisk data (ADASYN) brukes før en bygger modellen. Det er en metode som brukes for å korrigere ubalans i et datasett. ADASYN korrigerer ubalansen ved å lage syntetisk data fra kategorier som er underrepresentert. I vårt tilfelle gjelder det alle økter der vi registrerte skade.

3.13 Statistikk

For å gjøre statistiske analyser ble modellene SVM, RF, DT og LR brukt. Deskriptive data er presentert med gjennomsnitt og standardavvik (STD). Resultatene fra

maskinlæringsmodellene er presentert med presisjon, sensitivitet, F1 score og AUK. Vi gjennomførte kun analyser på data som ble inkludert i studien (Figur 8). Under oppsett av maskinlæringsmodellene har Java blitt brukt til å rense CSV filene. Scikit-learn er programmet som har blitt brukt til oppsett av de ulike modellene. Dette er et maskinlærings bibliotek for Python. I tillegg ble Python bibliotekene: Pandas, NumPy, Imblearn, Matplotlib og Seaborn brukt til datavisualisering, databehandling og rapportering.

3.14 Etikk

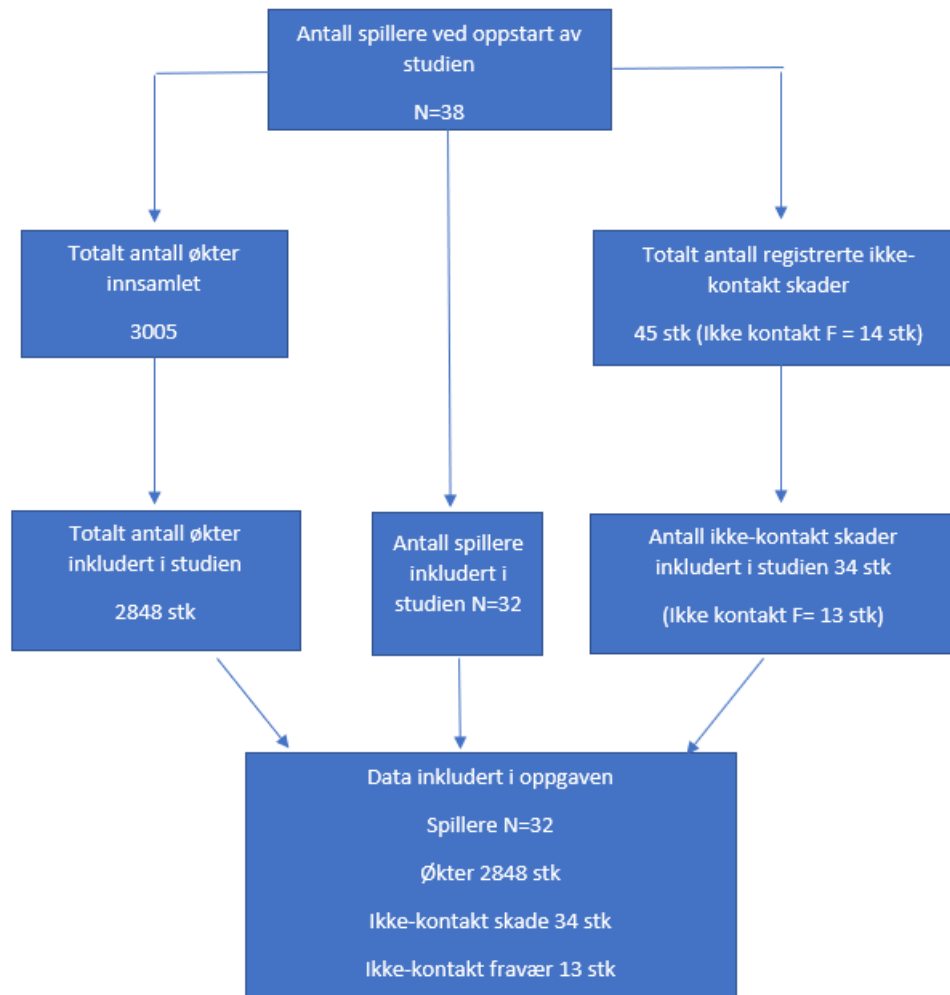
Alle spillerne som deltok i studien skrev under på et samtykkeskjema og ble informert om eventuelle risikoer med å delta. Det ble også gjort oppmerksomme på at deltagelse i studien var valgfritt, og at alle forsøkspersoner stod selv fritt til å trekke seg når som helst under studien uten å oppgi grunn for hvorfor de trakk seg. Studien er godkjent av Norsk senter for forskningsdata (Referansenummer 722773) og vurdert av NIH-REK. Alle forsøkspersonene vil bli aidentifisert og identifikasjonsnøkkel vil bli lagret i henhold til NIHs reglement, dvs. inntil fem år for etterprøvnbarhet etter prosjektslutt. Etter fem år vil nøkkelen bli destruert og alle forsøkspersonene forblir anonyme. Jeg som forsker påvirket ikke treningene eller kampene, og derfor var det ikke en økt risiko for skade, sykdom eller andre lignende faktorer. Jeg signerte en taushetsavtale med klubben, og i avtalen står det at jeg ikke kan dele informasjon om skader eller fotballtaktikk med andre utenfor klubben, utover det som står i masteroppgaven.

3.15 Budsjett og samarbeid

Studien er finansiert av Norges Idrettshøgskole, og i hovedsak er det reiseutgifter til og fra fotballklubben som er dekket. Det var et samarbeidsprosjekt med Universitetet i Oslo, som bisto med oppsett og analyse av data. Figur 11 og 12 er laget av samarbeidspartneren min, og figurene er laget i programmet han bruker til maskinlæringsmodellene. Derfor samsvarer ikke de to figurene med uttrykket til de andre figurene og tabellene i resultatkapittelet.

4. Resultater

Resultatene fra maskinlæringsmodellene baserer seg på data fra fotballsesongen 2019 i Norge. Utvalget består av 32 spillere fra eliteserien (Figur 8).



Figur 8: Flyttdiagram som viser antall spillere, treningsdata og antall skader som ble inkludert i prosjektet. Forkortelser: Ikke-kontakt F= Ikke kontakt skade som fører til fravær.



Figur 9: Figuren viser forholdet mellom totalt antall økter, og økter med registrert skade.

4.1 Skader

I de 11 månedene med datainnsamling registrerte vi totalt 34 IK skader. Av disse skadene ble 13 definert som IK-F. De resterende 21 ble definert som IK skade. Av en spillergruppe på 32 spillere var det 16 spillere som registrerte minimum en skade i løpet av sesongen. Det tilsvarer 1,06 (\pm 1,39) skader per spiller gjennom sesongen. Vi måtte fjerne 11 registrerte skader, disse skadene tilhørte spillere som ble ekskludert grunnet mangelfulle treningsbelastningsdata.

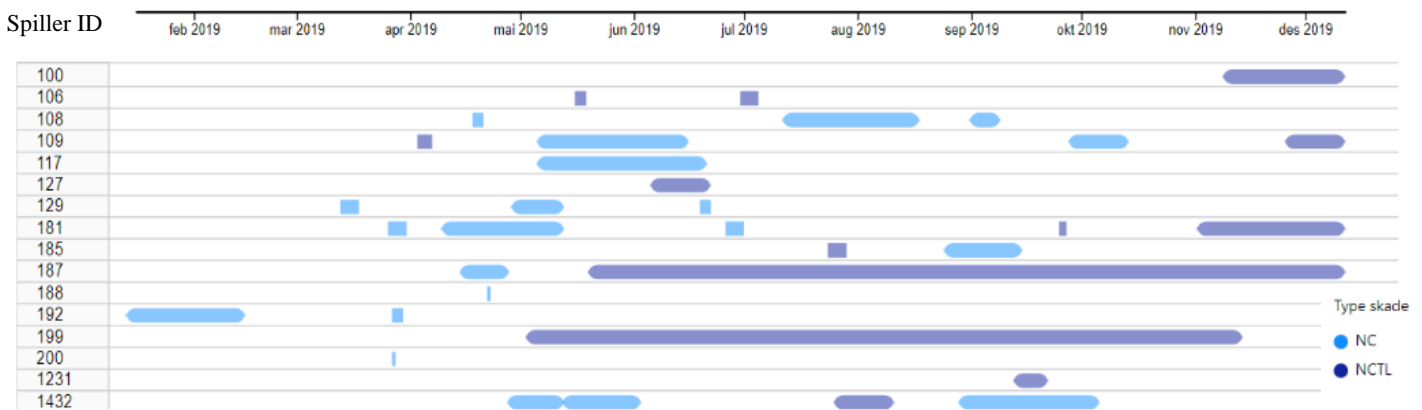
Tabell 2: Tabellen viser skadedata for begge skadedefinisjonene og totalt antall dager spillerne hadde fravær fra skade gjennom sesongen.

Total	Ikke-kontakt (IK)	Ikke-kontakt fravær (IK-F)
Total	21	13
Antall spillere	11	10
Spillere med 1 skade	5	7
Spillere med 2 skader	2	3
Spillere med 3 skader	4	-
Antall dager mistet i løpet av sesongen	-	534

Av alle skadene var 27 (79,4 %) i underekstremiteten. Av disse skadene var 19 (70%) akutte belastningsskader, og 8 (30 %) belastningsskader. De kroppsdelene hvor vi registrerte flest skader var kne (15,5 %), ankel (25 %) og hamstring (21,8 %). Vi registrerte flest akutte belastningsskader for ankel og hamstrings (15,6 %).

Tabell 3: Tabellen viser antall registrerte skader fordelt på lokalisasjon og antallet akutte belastningsskader og belastningsskader.

Kroppsdel	Registrerte skader (n)	Akutte belastningsskader (n)	Belastningsskader (n)
Kne	5	4	1
Ankel	8	5	3
Hamstring	7	5	2
Nedre del av rygg	2	0	2
Hofte	2	1	1
Forside lår	1	0	1
Legg	5	4	1
Fotblad	1	1	0
Mage	1	0	1
Hode	1	1	0
Bryst/ribbein	1	1	0
Total	34	22	12



Figur 10: Oversikt over tidspunkt for skadene og skadenes varighet gjennom 2019 sesongen. Søylen for IK-F skadene viser både varigheten for spillerens skade, samt antall dager spilleren hadde fravær fra trening eller kamp. IK-skadene er markert som varigheten på skaden, og ikke antall økter med fravær. Grunnen er at IK-skade per definisjon ikke forårsaker fravær fra trening eller kamp. Forkortelser: IK-skade= Ikke kontakt skade. IK-F = Ikke kontakt skade som fører til fravær.

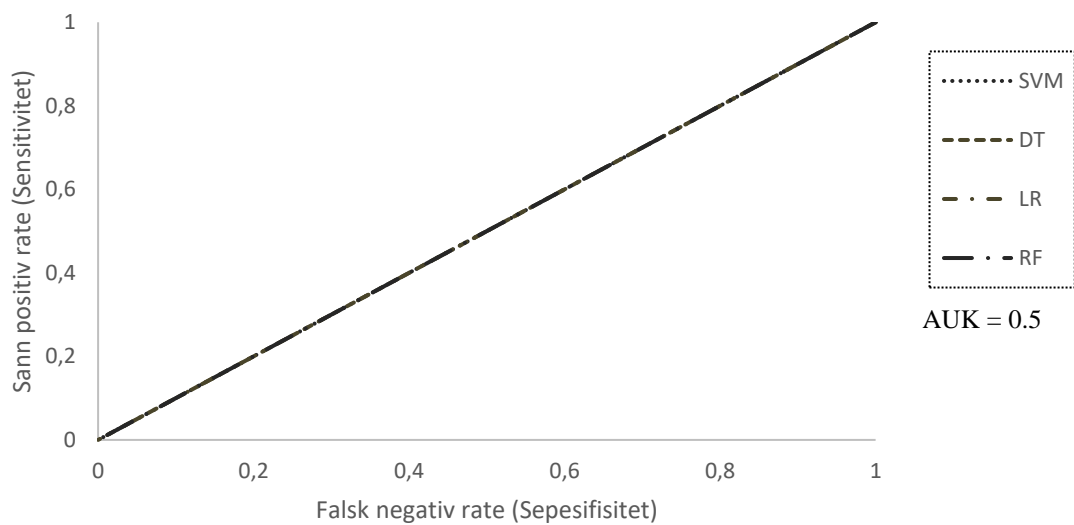
Figur 10 viser oversikt over skadene til spillerne gjennom 2019 sesongen i et Gantt chart. IK-skadene er markert med lyseblå og IK-F skadene med lilla farge. Hver markering i figuren viser datoen skaden ble registrert samt varigheten for skaden for en gitt spiller. IK-skadene er mest utbredt i første halvdel og i midten av sesongen. IK-F skadene er spredt utover hele sesongen, med 6 av skadene før juli, og 7 skader etter juli. To av IK-F skadene hadde en varighet på rundt 7 måneder, som tilsvarer over en halv sesong.

Resultatene fra tabell 4 er presentert med gjennomsnitt og standardavvik for 1000 tester av maskinlæringsmodellene. Sensitivitetsresultatene var det samme for alle modellene og viser 0.01 og 0.02 ± 0.01 . Det betyr at ingen av modellene klarte å predikere skade.

AUK scoren var den samme for alle modellene (figur 10), og ga et resultat som ikke var bedre enn ren gjetting.

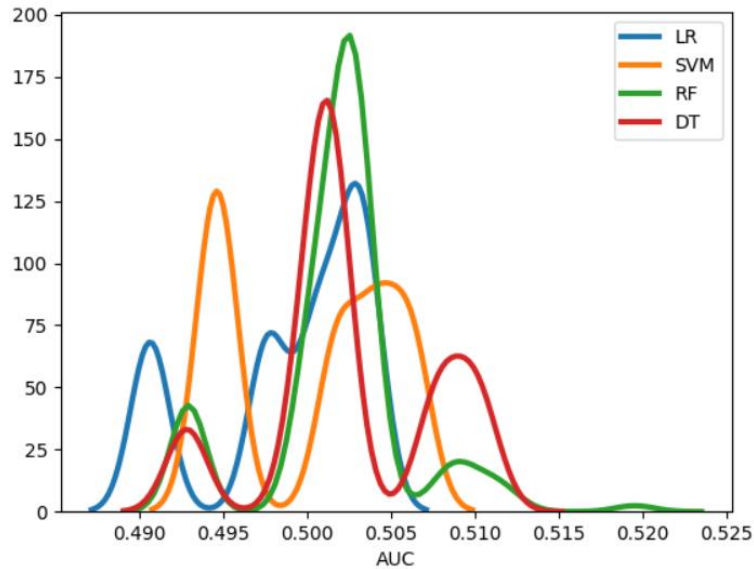
Tabell 4: Tabellen viser de fire forskjellige maskinlæringsmodellenes evne til å predikere skade for «ikke-kontakt skade» og «ikke-kontakt skade fravær», presenter med presisjon, sensitivitet, F1 og AUK score. Forkortelser: AUK= Areal under kurven.

Type maskinlæring	Type skade	Presisjon	Sensitivitet	F1-score	AUK
Support vector maskin	IK	0.45 ± 0.24	0.01 ± 0.01	0.02 ± 0.01	0.50 ± 0.01
	IK-F	0.45 ± 0.12	0.00	0.01	0.50
Logistisk regresjon	IK	0.33 ± 0.18	0.01 ± 0.01	0.02 ± 0.01	0.50 ± 0.01
	IK-F	0.46 ± 0.04	0.01	0.01	0.50
Random forest	IK	0.15 ± 0.07	0.02 ± 0.01	0.02 ± 0.01	0.50 ± 0.00
	IK-F	0.17 ± 0.08	0.01 ± 0.01	0.01 ± 0.01	0.50
Decision tree	IK	0.17 ± 0.08	0.02 ± 0.01	0.03 ± 0.03	0.50 ± 0.01
	IK-F	0.19 ± 0.10	0.01 ± 0.01	0.02 ± 0.01	0.50

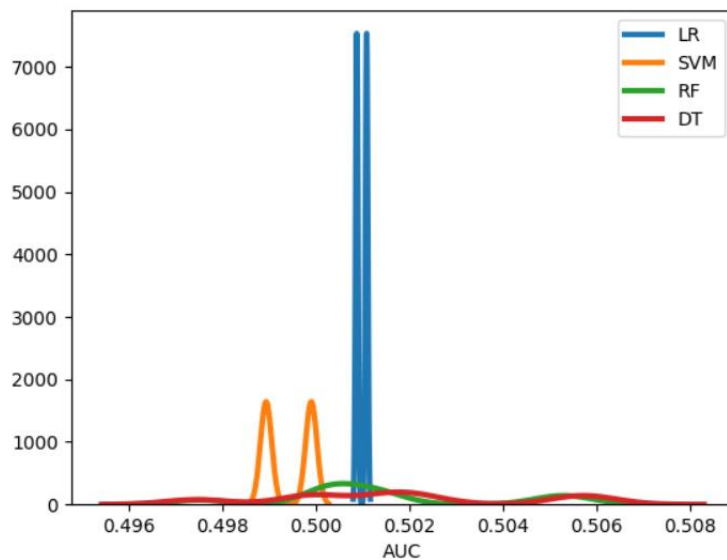


Figur 11: ROC kurve som viser arealet under kurven på 0.5 for maskinlæringsmodellen. Alle modellen fikk lik AUK. X-aksen representerer tilfeller der modellen har predikert riktig. Y-aksen viser tilfeller hvor modellen har predikert feil. AUK= Areal under kurven. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest

I figur 11 ser en at alle de fire forskjellige maskinlæringsmodellene vi brukte fikk en AUK på 0,5. Sann positiv rate illustrerer alle gangene modellen predikerte riktig (Eks: predikerer en skade som faktisk var en skade). En AUK på under 0,5 indikerer at modellen predikerer feil (Eks: Modellen predikerer at en skade, ikke er en skade). I vårt tilfelle klarer ikke modellen predikere en sammenheng mellom skade og treningsbelastning og modellen ender opp med å tippe hva som er skade og hva som ikke er en skade.



Figur 12: Tettheten til AUK for IK-skade ved 1000 gjennomkjøringer av algoritmen. Forkortelser: AUK= arealet under kurven. IK= Ikke kontakt skader. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest



Figur 13: Tettheten til AUK for IK-F skade fravær ved 1000 gjennomkjøringer av modellene. Forkortelser: AUK= Arealet under kurven. IK-F = Ikke kontakt fravær. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest

Figur 12 og 13 illustrer AUK-scoren for alle gangene vi har kjørt algoritmen for IK-skade og IK-F skade. Gjennomsnittet fra alle disse målingene blir 0,5 på alle de fire

modellene. X-aksen illustrerer tetthet og kan ikke direkte overføres til antall ganger modellen har blitt testet. Antallet på x-aksen er ikke det samme for figur 12 og 13 selv om vi testet den 1000 ganger på hver skadedefinisjon. Grunnen til dette er at modellen kompensere for liten range, som vi får på y-aksen. Som en ser av figur 13, så har alle AUK-scorene havnet mellom 0,495 til 5,08. Plottene viser at modellen er veldig usikker og ikke klarer å predikere sammenhengen mellom treningsbelastning og skade.

Gjennomsnitt og standardavvik i tabell 5 er regnet ut fra totalen av alle økter vi har med i studien (2848 økter). Vi hentet GPS-data fra 192 ulike dager gjennom 2019 sesongen, 35 av disse dagene inneholdt kamp og 157 av dagene var treningsøkter. I gjennomsnitt registrerte vi 89 økter per spiller totalt gjennom sesongen.

Tabell 5: Tabellen viser gjennomsnitt og standardavvik for distanse, playerloadTM, acceleration load, V4 distanse, V5 distanse og høyhastighetsløp. (Gjennomsnitt er beregnet fra hele sesongen).

Parameter	Gjennomsnitt	Standardavvik
Distanse	5095 m	± 834,83 m
Player load TM	521,62	± 103,69
Acceleration load	1550,93	± 206,95
V4 distanse	186,61 m	± 68,66 m
V5 distanse	54,98 m	± 27,31 m
Høyhastighetsløp	241,60 m	± 93,10 m

5. Diskusjon

Formålet med denne studien var å se om vi med hjelp av multidimensjonale modeller klarte å predikere sammenhengen mellom IK-skader og treningsbelastning i et norsk mannlig eliteserielag. Hovedfunnene i studien er at modellen og variablene vi brukte for å predikere skade basert på høyhastighetsløp (HSR^{EWMA}), total distanse ($Dist^{MSWR}$) og tidligere skade (PI^{EWMA}) ikke klarte å predikere ny skade. Gjennomsnittlig AUK for alle modellene ble målt til 0,5, som indikerer at prediksjonen ikke er bedre enn tilfeldig gjetting. Presisjonsnøyaktigheten var høyere enn sensitiviteten på alle modellene. LR er modellen som målte den høyeste presisjonen, 0.46 ± 0.04 . Ingen av modellene fikk en høyere sensitivitet enn 0,02, som indikerer at modellen heller ikke er i stand til å identifisere type skade.

5.1 Prediksjon av skade i fotball

Å predikere skade i fotball er ingen lett oppgave, det er et av de mest utfordrende problemene innen idrett, samtidig som det er en nøkkelfaktor for forebygging av skade (Bittencourt, 2016). Kroppen til et menneske er komplisert, og skade kan være forårsaket av utallige elementer som skjer både på og utenfor banen. Det har blitt diskutert at ved å implementere en mer kompleks tilnærming til idrettsskadeforskning, vil det bidra til å bedre forstå og forhindre skade (Bittencourt et al., 2016; Nielsen, Malisoux, Møller, Theisen, & Parner, 2016). På grunn av skaders multifaktorielle natur så er det usannsynlig at data fra kun en tilnærming (GPS-data), er nok til å få prediktiv «power» til å identifisere skader (Bahr, 2016).

Ulike belastningsparameter (tabell 5) har vært mye brukt i praksisfeltet for å se om det kan bidra til skadeprediksjon, fordi de er modifiserbare faktorer. Utfordringen med å predikere skade er at flere risikofaktorer knyttet til skade er ikke-modifiserbare (Eckard et al., 2018). Som tidligere nevnt er dette faktorer som alder, kjønn, skadehistorikk og konkurransenivå (Hägglund et al., 2006; Murphy, Connolly, & Beynnon, 2003). Ikke-modifiserbare faktorer går det ikke an å endre på, og dermed blir det naturlig at det testes om modifiserbare faktorer som treningsbelastning potensielt kan bidra til å predikere skade.

Resultatene fra studier som har prøvd å predikere skade ved hjelp av maskinlæringsmodeller varierer betydelig. Dette kan delvis forklares med de metodiske ulikhetene studiene har. Studien til Carey et al. (2018) bruker tre av de samme maskinlæringsmodellene vi brukte, den studien så på prediksjon av IK-skader og hamstringskader basert på treningsbelastning. Carey et al. (2018) rapporterte at de multivariate modellene var marginalt bedre enn de univariate modellene på prediksjon av IK-skader og IK-F. Samtidig viste modellene deres en begrenset evne til å predikere IK-skader, mens de derimot fikk bedre resultater når det gjaldt prediksjon av hamstringsskader (AUK = 0,76). Det indikerer at en mer spesifikk skadedefinisjon gir bedre prediksjon. Modellen til Carey et al. (2018) viste en tendens til forbedring av resultatene når mengden data økte, selv om konklusjonen fra forfatterne var at modellen viste dårlig evne til å predikere skade. Resultatene til Carey et al. (2018) er sammenfallende med resultatene fra vår studie, og viser begrensingen ved å kun inkludere data fra et lag over en sesong.

En annen studie som er gjennomført i bare en sesong, er Alessio Rossi (2018). Rossi fikk resultater med en god nøyaktighet, og de klarte å predikere 80 % av IK-skader med 50 % presisjon ved bruk av DT. Vi brukte samme metode som Rossi, og det er derfor overraskende at våre data gir motstridende resultater, selv med en større datamengde. Ettersom begge studiene er gjennomført med data basert på ett lag fra en sesong, så ville det vært naturlig at den studien med det største datasettet ville predikert skader med bedre nøyaktighet. Dette er ikke tilfelle, og det kan derfor stilles spørsmålsteget ved modellens generaliserbarhet.

Det kan være flere årsaker til at vi ikke får tilsvarende resultater som studien til Alessio Rossi (2018). Det første jeg vil trekke frem er at vi i vår studie har et ubalansert datasett (figur 9). Optimalt sett skal en ha like mange økter uten registrert skade som med registrerte skader. Da har man et perfekt balansert datasett. Ubalanserte datasett gjør det vanskeligere for maskinlæringsmodellens evne til å lære og SVM er en modell som viser seg å være ineffektiv på ubalanserte datasett (Gareth James., 2014).

Maskinlæringsmodellen fokuserer på den vanligste hendelsen (økter uten skader) og ender opp med å ignorere den sjeldne hendelsen (økter med skader) (Menardi & Torelli, 2014). For å balansere datasettet har vi brukt ADASYN. En mulig svakhet ved bruk av ADASYN er at de syntetiske dataene blir for like. Siden ADASYN baserer seg på de 34

skadene som ligger i datasettet blir det vanskelig for ADASYN å måtte lage mange syntetiske skadedata.

Verghese et al. (2018) påpekte i sin studie at datasettet som brukes i modellen er den viktigste delen i prosessen. Siden vi valgte å bruke samme parametere og maskinlæringsmodeller som Alessio Rossi (2018), så kunne vi bruke all dataen vår som testdata, vi slapp på den måte å bruke en del av dataene våre som treningsdata for lære opp maskinlæringsmodellen. Dobbin og Simon (2011) hevder at det ikke er noe konsensus på hvor mye data som bør brukes til å trene opp modellen, men i deres beregninger fant en at ved bruk av et sted mellom 40-80 % av dataene ville resultere i optimal prediktiv evne. At vi i våre analyser kunne bruke 100 % av dataene som testdata burde være positivt og kan sees på som en styrke ved studien siden vi da fikk med alle skadene.

Å teste en maskinlæringsmodell på et ukjent datasett er nødvendig for å finne modellens evne til å predikere. Maskinlæringsmodeller har mange frihetsgrader og kan spesialisere seg på et datasett slik at den klarer å predikere tilnærmet 100 %. Dette kalles «overfitting», og er noe en ønsker å unngå. For å teste om modellen er generaliserbar og kan brukes, må man teste modellen på et nytt og ukjent datasett (Max Kuhn, 2013). I denne studien unngikk vi overfitting ved at vi testet modellen på hele datasettet med en gang. Dette kunne vi gjøre siden modellen vår er bygd opp likt som Alessio Rossi (2018). Modellen til Rossi er en ferdig modell, det gjør at vi ikke trengte å trene opp modellen før vi testete den.

Formålet til maskinlæringsmodellene er å gi trenere et praktisk verktøy som kan brukes til å ta valg om endringer i treningsbelastning eller kampbelastning for å forhindre skader på spillere (Kirkendall & Dvorak, 2010). Disse utfordringene er knyttet til modellens nøyaktighet og tolkning av data (Alessio Rossi, 2018). Prediksjonsmodeller bør være lette å tolke basert på at trenere må kunne vite *hvorfor* en spiller er i risikozonen for en skade. Dette gjør at en trener kan gjøre de endringer som er nødvendige for å unngå situasjoner som medfører høy risiko for skade. DT blir sett på som en maskinlæringsmetode som er enkelt å tolke, men en svakhet er metodens prediksjonsnøyaktighet sammenlignet med andre typer maskinlæringsmodeller (Gareth James., 2014). At en prediksjonsmodell er nøyaktig er svært viktig, da «falske alarmer» kan ha en negativ innvirkning på måten trenere, basert på resultater fra modellen, velger å legge opp treningsbelastningen til en gitt spiller eller gruppe spillere.

5.2 Er maskinlæring basert på GPS-data komplekst nok?

Flere studier har prøvd seg på en mer kompleks fremgangsmåte (Alessio Rossi, 2018; Carey et al., 2018; Ruddy et al., 2019). Selv om maskinlæring har økt i popularitet, er det fortsatt begrenset med studier på området (Claudino et al., 2019). Biomekaniske-, fysiologiske- og psykologiske forhold mener Bittencourt et al. (2016) er viktig å ta med hvis en skal klare å predikere skade, med tanke på at skade oppstår fra ikke-lineære sammenhenger og fra multifaktorielle forhold. I denne studien har vi kun sett på fysiologiske faktorer gjennom GPS-data, og det kan ses på som en svakhet at studien ikke ha med data på biomekaniske eller psykologiske forhold. Sett i forhold til Bittencourt et al. (2016) studie så er denne studiens tilnærming en forenkling på et komplekst problem. En risikofaktor (eks. tidligere skade) er ingen garanti for at det kommer til å oppstå en ny skade (Bittencourt et al., 2016). På den andre siden så kan identifisering av en risikofaktor gi informasjon om sannsynligheten for en ny skade. Skade i idrett er et komplekst problem med flere usikkerhetsmomenter, og med bakgrunn i dette mener Bittencourt et al. (2016) at den beste måten å predikere skade er å prøve å forstå sammenhengene mellom de ulike faktorene, og ikke se på hver av faktorene for seg. Vi valgte på forhånd ut tre parameter (Høyhastighetsløp HSR^{EWMA} , total distanse $Dist^{MSWR}$, og tidligere skade PI^{EWMA}). Basert på modellen til Bittencourt et al. (2016) kan det se ut som denne studien har fokusert for mye på de enkelte faktorene, og med den metoden som er brukt her ikke har klart å se på hele kompleksiteten av ulike sammenhenger som kan ligge i et datasett. For at vi skulle fått en mer fullstendig oversikt over de enkelte faktorene, så anbefales det at man i tillegg til GPS-data blant annet samler inn data på stress, angst og mestringsevne ved hjelp av et selvrappoteringskjema, og i tillegg også data på søvnkvalitet og ernæring. Dessuten kan og bør det gjennomføres mobilitetstesting og testing av stabilitet, kraft og motorisk kontroll (Stern et al., 2020).

Å predikere skade basert på spillernes trenings- og kampbelastning, er ikke eneste metoden man kan bruke til å predikere. I studien til Rommers et al. (2020) tok de antropometriske målinger (høyde, vekt), motoriske- og fysiske tester (Fleksibilitet, utholdenhet, hastighet, agility) før sesongen startet. De brukte en XGBoost (Maskinlæringsmetode) til å predikere skade gjennom sesongen basert på resultatene fra baseline testene før sesongstart. Rommers et al. (2020) konkluderte med at modellen predikerte ganske nøyaktig skader basert på testresultatene før sesongen. Dette er en

forholdsvis ny studie, og det kunne vært interessant og inkludert GPS-data i modellen. Fordelen med GPS-data er at en hele tiden får nye tall inn i modellen som gjør at man kan holde kontroll på treningsbelastningen. Ifølge Blanch og Gabbett (2016) er jevn og gradvis økende treningsbelastning viktig for å unngå økt skaderisiko.

En større mengde data vil kunne forbedre modellens evne til finne sammenhenger i datasettet og på den måten øke prediksjonsevnen. Det er vanskelig å estimere mengden data vi trenger for å få en mer presis modell, og det er ikke sikkert vi hadde fått noe annerledes resultater selv med data fra 10 sesonger. En studie fra Peduzzi et al. (1996) foreslo at man trenger minst 10 skader per kroppsdel for å klare å predikere sammenhenger. Bahr og Holme (2003) anbefaler at man bør ha 20 til 50 skader for å finne en moderat til sterk assosiasjon mellom skade og risikofaktorene. Det lave antall skader er en svakhet i denne studien, men samtidig er muligheten til å samle inn store skadedatataill en utfordring og muligens begrensning innen idrettsmedisinsk forskning (van Dyk et al., 2018).

5.3 Assosiasjon og prediksjon

Det er viktig å skille mellom assosiasjon og prediksjon. To variabler er assosiert når den ene variabelen gir informasjon om den andre (Ruddy et al., 2019). Hvis en variabel forårsaker et spesifikt utfall hos en annen variabel, er de to variablene direkte assosiert. Studier som ser på assosiasjoner, er viktig fordi de kan bidra til å identifisere risikofaktorer som igjen kan gi viktig informasjon om skadeetiologien. Slike studier kan hjelpe oss å forstå hvorfor skader oppstår (McCall, Fanchini, & Coutts, 2017). To variabler kan ha en assosiasjon, selv om det ikke forårsaket et direkte utfall (Altman & Krzywinski, 2015). Indirekte assosiasjon oppstår når en variabel påvirker utfallet av en variabel via en mellom variabel (Altman & Krzywinski, 2015). Det er også mulighet for at en assosiasjon mellom to variabler er tilfeldig, og ikke er forårsaket av noe. Dette refereres til som en type 1 feil. En type 1 feil oppstår når en assosiasjon/effekt av noe blir funnet, men i realiteten finns det ikke noe assosiasjon/effekt. Type 2 feil oppstår når en ikke finner en assosiasjon/effekt, men i virkeligheten er det en assosiasjon/effekt (Smoliga & Zavorsky, 2017).

Prediksjon er evnene til å anvende datagrunnlaget en har samlet inn til å forutsi, basert på kjente verdier, til ukjente utfall (Shmueli, 2010). Å forstå nyansene mellom

assosiasjon og prediksjon kan være vanskelig. Det kan resultere i at en trener konkluderer med at en faktor som er forbundet med skade kan brukes til å predikere skade og på den måten forhindre skade (McCall et al., 2017). I sammenheng med skader kan ulike assosiasjoner hjelpe oss å identifisere ulike brikker i puslespillet om hvorfor skader oppstår, mens prediksjon har evnene til å bruke hele puslespillet til å komme med forutsigelser (Shmueli, 2010). Et eksempel på en studie som har prøvd å predikere skade basert på etablerte risikofaktorer er studien til Ruddy et al. (2019). Ruddy et al. (2019) implementerte en maskinlæringsmodell og prøvde å predikere skade basert på alder, tidligere hamstringsskade og eksentrisk hamstring styrke. De konkluderte med at på tross av at disse faktorene er assosiert med hamstring skader, kunne man ikke bruke de til å predikere skader (Ruddy et al., 2019).

5.4 Skadelokalisasjon og skademekanismer

I denne studien var 70 % av skadene akutte belastningsskader, mens de resterende 30 % skadene var belastningsskader, uten et klart skadetidspunkt. Faude et al. (2013) rapportert at i gjennomsnitt er 60-90% av skadene i fotball er akutte skader, mens 10-40% er klassifisert som overbelastningsskader. Figur 9 gir en god indikasjon på når skadene oppsto igjennom sesongen. Naturlig nok registrerte vi flest IK-skader, med tanke på definisjonen til IK-F også inneholder at spilleren må være borte fra minimum en trening eller kamp. To av IK-F skadene hadde en varighet på nesten en halv sesong, og dette påvirket totalt antall dager borte fra trening eller kamp i betydelig grad. Som et resultat av fraværet til disse spillerne, har de begrenset med GPS-data. Dette er en faktor som kan ha påvirket modellenes evne til å predikere.

5.5 Valg av skadedefinisjon

Det er gjennomført mange epidemiologiske studier av skader i fotball, og valget av skadedefinisjon varierer i stor grad. Konsekvensen er at det kan være krevende å sammenligne resultater og konklusjoner. Med bakgrunn i dette etablerte det internasjonale fotballforbundet (FIFA) i 2005 (Fuller et al., 2006) en gruppe bestående av erfarne forskere for å lage enhetlige skadedefinisjoner som kunne brukes i epidemiologiske studier i fotball og følgende skadedefinisjoner ble beskrevet; *“Any physical complaint sustained by a player that results from a football match or football training, irrespective of the need for medical attention or time-loss from football. An*

injury that results in a player receiving medical attention is referred to as a “medical-attention” injury and an injury that results in a player being unable to take a full part in future football training or match play as a “time-loss” injury” (Fuller et al., 2006).

Det vil si at det er tre ulike alternativer for definisjon av en skade. (1) fysiske plager, (2) medisinsk tilsyn, (3) skade som fører til fravær. Selv om forskergruppen anbefaler at man bruker alle de tre definisjonene, er skade som fører til fravær mest brukt.

(Bjørneboe et al., 2016; Ekstrand et al., 2011b). Skade som fører til fravær er den mest spesifikke og smaleste definisjon, og er den definisjonen som mest sannsynlig rapporterer minst insidens (Clarsen, 2014). En svakhet med «skade som fører til fravær» definisjonen er at grensen for hva som skal til for å rapportere skade varierer mellom forskjellige støtteapparat og spillere. Det er mindre sannsynlig at en sentral spiller vil være borte fra kamp eller trening på grunn av en liten skade under en viktig del av sesongen, enn mindre viktige spillere vil være (Clarsen, 2014). Rapporteringen av «medisinsk tilsyn» skader er anbefalt hos flesteparten av konsensus anbefalingene fordi den mest sannsynlig fanger opp flere tilfeller av skader, og vil derfor gi et bedre bilde av antall skade og sykdom (Hodgson, Gissane, Gabbett, & King, 2007). Det har vært et kjent problem innen idrettsskadeepidemiologien at det er vanskelig å registrere belastningsskader på en valid og reliabel måte. Clarsen et al. (2013) utviklet et nytt spørreskjema for registrering av belastningsskader i idrett. Hensikten var å kunne fange opp alle belastningsskader uavhengig av kroppsdel, med bakgrunn av at utøvere ofte fortsetter å trene/konkurrere til tross for eksistens av belastningsskader (Clarsen et al., 2013). Utøverne måtte svare på spørreskjemaet ukentlig i 13 uker. I studien sammenligner de rapportering av belastningsskader ved bruk av spørreskjema, og den mer klassiske metoden hvor det er fysioterapeut eller trener som rapporterer «skader som fører til fravær». Resultatene til Clarsen et al. (2013) viser at det er mer enn 10-ganger så mange plager knyttet til overbelastning, enn «skade som fører til fravær».

Valg av skadedefinisjon vil påvirke hvor mange skader man får rapportert gjennom en sesong. Hulin (2017) påpeker at studier som har sett på treningsbelastning har vært inkonsekvente når det kommer til valg av skadedefinisjon. Hulin (2017) poengterer at hvis man ser bort i fra skadedefinisjon når en analyserer resultatene, kan dette påvirke funnene. Han refererer til forskning som viser forskjellig resultat ved bruk av forskjellige skadedefinisjoner (Finch, 1997; Hulin, 2017; Windt & Gabbett, 2017). Det er ikke enighet om hvordan belastningsskader skal defineres (Bahr, 2009). En standard

skadedefinisjon er viktig ved sammenligning av studier. I fotball har det vært vanlig å bruke IK-F definisjon på studier som ser på belastningsskader (Hulin, 2017; Windt & Gabbett, 2017). Det kan være vanskelig å definere hva som er en belastningsskade, spesielt hvis belastningsskaden opptrer med akutte symptomer. Da kan det bli kategorisert som en akuttskade, men skaden skyldes overbelastning (Bahr, 2009). Valget av skadedefinisjon er forskjellig mellom studier, noen inkluderer kontaktskader, andre bruker ikke-kontaktskader i analysen sin (Jaspers et al., 2018; Malone et al., 2017b). Flere studier har brukt en definisjon som forutsetter at spilleren ikke trener eller spiller kamper i en gitt tidsperiode for at det skal registreres som en sykdom/skade (Bowen, Gross, Gimpel, Bruce-Low, & Li, 2019; Bowen, Gross, Gimpel, & Li, 2017; Lathlean, Gastin, Newstead, & Finch, 2019). Utfordringen med denne definisjonen er at den ikke fanger opp alle belastningsskader, da belastningsskader ikke alltid fører til tid borte fra trening eller kamp. En styrke i denne studien er at vi inkluderte to ulike definisjoner på skade. Dette er spesielt viktig når en studerer belastningsskader, som ikke nødvendigvis fører til fravær fra trening eller kamp. Dette kan være en mer relevant skadedefinisjon når en studerer forholdet mellom treningsbelastning og skader. Hadde vi kun brukt IK-F definisjonen i vår studie, så ville vi fanget opp 13 av 34 skader. Det tilsvarer 44 % av skadene. Bruken av ulike skadedefinisjoner i studien vår forbedret ikke prediksjonsevnen til modellene, dette resultatet støttes av en lignende studie gjort på Australsk fotball (Carey et al., 2018).

5.5.1 Valg av monitoreringsmetode

Flere forskjellige metoder kan brukes for å samle inn data om intern- og ekstern treningsbelastning. Det finnes ingen metoder der man klarer å inkludere alle aspekter ved treningsbelastning (Soligard et al., 2016). Når man studerer forholdet mellom treningsbelastning og skader, vil valget av metode kunne påvirke resultatet, fordi de ulike metodene måler forskjellige aspekter ved belastning. Forskjellige monitoreringsmetoder kan i noen tilfeller egne seg bedre til enkelte skader. For eksempel vil bruk av sRPE ikke være en ideell metode til å se på forskjeller i sprintintensitet, en faktor som kan ha en betydning for skaderisikoen for hamstringsskader (Buchheit, 2017). Å bruke GPS for å måle antall høyintensitetsløp vil være en bedre og mer egnet metode hvis man eksempelvis ser på hamstringsskader.

Etter analyse av GPS-dataen vår fant vi at flere GPS-brikker hadde store feilmålinger. Dette gjorde at noen av spilleren fikk urealistiske tall i noen perioder og førte til at disse spillerne ble ekskludert fra datagrunnlaget. Slike feilmålinger kan ha skjedd uten at vi har lagt merke til det. Det kan ha påvirket resultatene våre, og vi må da stille spørsmål til relabiliteten til GPS-enhetene. Rampinini et al. (2015) fant at validiteten til brikkene ble svekket ved høyere hastighet (Variasjonskoeffisient 10,5%). Videre fant Akenhead, French, Thompson og Hayes (2014) i sine studier at validiteten til GPS-enhetene var begrenset ved akselerasjoner og korte sprinter. Dette er interessante funn, som betyr at flere av feilmålingene kan ha forekommet i perioder med enten høye hastigheter eller spill på små flater. GPS-dataen vi brukte kan også være en mulig begrensing. Vi oppdaget i noen tilfeller GPS-brikker gikk tom for strøm midt i treningene. For at disse spillerne kunne være med i studien måtte vi lage et gjennomsnitt fra resten av laget på den gitte treningen til den spilleren hvor GPS-brikken skrudde seg av. Dette er en svakhet med tanke på at den registrerte GPS-dataen til spilleren ikke er de sanne verdiene til spilleren.

I denne studien inkluderte vi bare ekstern treningsbelastning. Intern treningsbelastning ble ikke tatt med, selv om det har blitt assosiert med skader i flere studier (Soligard et al., 2016). Ved innsamling av treningsbelastningsdata, vil en kombinasjon av intern- og eksterne belastning være å anbefale (Burgess, 2017), og vi burde tatt med andre treningsaktiviteter som styrketrening og spesifikk rehabilitering for å kunne ha et enda bedre bilde på den totale belastningen til spillerne.

5.6 Begrensinger i denne studien

Først av alt er det viktig å understreke at det var masterstudentene som stod for de sentrale rollene som utviklere av maskinlæringsmodellen. Ved oppstarten av dette prosjektet så hadde ingen av masterstudentene noe særlig kunnskap rundt maskinlæring. Med den forkunnskapen som lå i bunn og at studien til Alessio Rossi (2018) hadde fått til å predikere skade, ble modellen utviklet med bakgrunn på dette. Det kan derfor hende at resultatene i denne oppgaven ble påvirket av feiltolkningen av metodebeskrivelsen til Alessio Rossi (2018).

Det er flere begrensninger som kan ha påvirket modellenes evne til å predikere IK-skade. Studien var en longitudinal kohortstudie av et fotballag. En generell ulempe med

kohortstudier er at det vil være tidkrevende å følge en kohort lenge nok til det har blitt registrert nok sykdomstilfeller/skadetilfeller til å kunne trekke en konklusjon (Laake, Olsen, & Benestad, 2008). Mengden data som ble samlet inn er mindre enn i flere andre studier som har gjort lignende, disse studiene har flere forsøkspersoner eller flere sesonger med data (Carey et al., 2018; Jaspers et al., 2018; Rommers et al., 2020; Ruddy et al., 2019). Med begrenset størrelse på datasettet kan det påvirke modellens evne til å generalisere skadene, og man ender opp med risiko for overfitting. Det er foreslått at man trenger data fra mer enn 10 sesonger for å lage en reliabel prediksjonsmodell (Carey et al., 2018). Med tanke på at studien kun inneholder et lag, er det en begrensning at vi ikke kan generalisere resultatene til andre lag, dette med bakgrunn i de taktiske, tekniske og individuelle forskjellene mellom lagene. Eliteserielaget som deltok i denne studien hadde en høy utskifting av spillerstallen midt i sesongen, noe som kan ha påvirket modellens nøyaktighet for prediksjon av skader. Det er to muligheter for å øke størrelsen på datasettet. Den ene måten er å inkludere flere lag, og det kan oppnås ved at lagene i eliteserien deler GPS-dataen sin. Hadde vi fått til dette ville et bredt utvalg av spillere og mer GPS-data forbedret generaliseringen til modellen. I tillegg kan en stor mengde data samles inn på kort tid. Den andre måten er å følge et lag over flere sesonger. Optimalt sett er en kombinasjon av disse to metodene ønskelig.

Få skader er en annen konsekvens av å gjennomføre denne studien på kun et lag gjennom en sesong. IK-F skade består av 0,5 prosent av hele datasett vårt. Sammenligner vi det med andre studier har de rapportert en insidens fra 1,7 til 2,2 prosent av datasettet (Alessio Rossi, 2018; Carey et al., 2018). Ved å inkludere flere klubber, og ved å fortsette å samle inn data på belastning og skader så vil antall innrapporterte skader øker.

Maskinlæringsmodellen fungerer slikt at jo mer data man kjører igjennom den, jo bedre blir den på å predikere. Forskning som inneholder maskinlærings, er komplisert og inneholder mye statistikk som kan være vanskelig å tolke. I noen av tilfeller gjetter modellen seg frem til resultatene og andre tilfeller trenger man en egen data for å tolke resultatene.

5.7 Videre forskning

I dette prosjektet har vi undersøkt om en maskinlæringsmodell klarer å predikere skade i en eliteserielubb i Norge. Det foreligger en del studier på området, og forskningsfeltet har fått en økning i popularitet. De prosjektene som har sett på fotballskader har i all hovedsak fokusert på mannlige elite utøvere. Flere studier trengs for å studere dette både blant kvinner og menn på elitenivå, og også på junior elitenivå. Dessuten bør dette gjøres over flere sesonger. Å teste maskinlæring som verktøy i andre lagidretter som innhenter data er også mulig.

Det viktigste for videre forskning er å samle inn større mengde data, både skadedata, men også andre data som tar hensyn til kompleksiteten når skader oppstår. Dette er avgjørende for å klare å kunne predikere skade i fremtiden. Data fra et større utvalg vil dessuten kunne forbedre generaliserbarheten til maskinlæringsmodellen.

Det er kjent at valget av skadedefinisjon vil påvirke antall rapporterte skader. Studier som skal se på maskinlæring og prediksjon av skade bør ta i betraktning at valg av skadedefinisjon vil ha en påvirkning på resultatet. Fremtidig forskning på området bør vurdere å bruke flere forskjellige skadefinisjoner og se om man klarer å predikere på en av dem isolert sett. Det bør også tas hensyn til hvilke typer skader en ønsker å predikere. I denne studien ble det brukt en bred definisjon, men det går også an å prøve på en mer spesifikk tilnærming på en skade. I vår studie brukte vi kun fysisk belastning i form av GPS-tall. Videre forskning bør tenkte på andre aspekter som psykologiske aspekter, samt andre tilleggsaspekter ved fysisk belastning. Det kan være data som hjertefrekvens fra treningene, laktat, kosthold, søvn, eller sRPE.

Maskinlæring åpner også opp andre potensielle muligheter innen fotball. Det kan for eksempel bli brukt til å identifisere spesifikke aktivitetsprofiler etter hvilken posisjon spilleren spiller under kamper, samt evaluere spesifikke kampsituasjoner. Man kan bruke det på yngre spillere for å finne talenter, basert på fysisk karakteristika og treningshistorikk slik som Gonçalves, Coelho e Silva, Carvalho, og Gonçalves (2011) gjorde i sin studie. Dataen fra dette prosjektet ble brukt til å predikere skader, men ved hjelp av maskinlæring kan muligens fremtidig forskning se på mulighet for å predikere når en spillere er tilbake fra en skade.

6. Konklusjon

De multidimensjonale modellene klarte ikke å predikere ikke-kontaktskader i profesjonell fotball basert på GPS-data fra en sesong. Resultatene i studien viser at bruk av to ulike skadedefinisjoner ikke forbedrer modellenes evne til å predikere skade.

Basert på innhentede data i denne studien, kan man ikke bruke modellen som et verktøy i treningshverdagen. GPS-dataen som ble brukt viser begrensningen med maskinlæring, men med hensyn til videre forskning vil inkludering av flere parameter av flere spillere over flere sesonger være fordelaktig. Samtidig kan det være fordelaktig å predikere en mer spesifikk skadetype. Ved at man tar i bruk komplekse tilnæringer kan vi komme et steg nærmere prediksjon og forståelse av bakenforliggende årsaker til skader. For å finne ut mer om en kompleks og multifaktoriell tilnærming er den rette veien videre, trengs det mer forskning på området.

Referanser

- Akenhead, R., French, D., Thompson, K. G., & Hayes, P. R. (2014). The acceleration dependent validity and reliability of 10Hz GPS. *Journal of Science and Medicine in Sport*, *17*(5), 562-566.
doi:<https://doi.org/10.1016/j.jsams.2013.08.005>
- Akenhead, R., & Nassis, G. P. (2016). Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *11*(5), 587.
doi:10.1123/ijsp.2015-0331
- Alessio Rossi, L. P., Paolo Cintia, F. Marcello Iaia, Javier Fernández, Daniel Medina. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*. doi:<https://doi.org/10.1371/journal.pone.0201264>
- Altman, N., & Krzywinski, M. (2015). Points of Significance: Association, correlation and causation. . *Nature methods*, *12*(10), 899-900.
doi:doi.org/10.1038/nmeth.3587
- Andersen, T. E., Tenga, A., Engebretsen, L., & Bahr, R. (2004). Video analysis of injuries and incidents in Norwegian professional football. *British Journal of Sports Medicine*, *38*(5), 626-631. doi:10.1136/bjism.2003.007955
- Arnason, A., Andersen, T. E., Holme, I., Engebretsen, L., & Bahr, R. (2008). Prevention of hamstring strains in elite soccer: an intervention study. *Scandinavian Journal of Medicine & Science in Sports*, *18*(1), 40-48. doi:10.1111/j.1600-0838.2006.00634.x
- Arnason, A., Sigurdsson, S. B., Gudmundsson, A., Holme, I., Engebretsen, L., & Bahr, R. (2004). Risk Factors for Injuries in Football. *The American Journal of Sports Medicine*, *32*(1_suppl), 5-16. doi:10.1177/0363546503258912

- Bahr, R. (2009). No injuries, but plenty of pain? On the methodology for recording overuse symptoms in sports. *British Journal of Sports Medicine*, 43(13), 966-972. doi:10.1136/bjism.2009.066936
- Bahr, R. (2016). Why screening tests to predict injury do not work—and probably never will...: a critical review. *British Journal of Sports Medicine*, 50(13), 776-780. doi:10.1136/bjsports-2016-096256
- Bahr, R., & Holme, I. (2003). Risk factors for sports injuries--a methodological approach. *British Journal of Sports Medicine*, 37(5), 384-392. doi:10.1136/bjism.37.5.384
- Bahr, R., & Krosshaug, T. (2005). Understanding injury mechanisms: a key component of preventing injuries in sport. *British Journal of Sports Medicine*, 39(6), 324-329. doi:10.1136/bjism.2005.018341
- Bahr, R., McCrory, P., Laprade, R., Meeuwisse, W., & Engebretsen, L. (2014). *Idrettsskader: diagnostikk og behandling* Fagbokforlaget.
- Bittencourt, N. F. N., Meeuwisse, W. H., Mendonça, L. D., Nettel-Aguirre, A., Ocarino, J. M., & Fonseca, S. T. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British Journal of Sports Medicine*, 50(21), 1309-1314. doi:10.1136/bjsports-2015-095850
- Bittencourt, W. H. M., Mendonça, A Nettel-Aguirre, Ocarino, Fonseca. (2016). Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. *British Journal of Sports Medicine*. doi:<http://dx.doi.org/10.1136/bjsports-2015-095850>
- Bjørneboe, J., Kristenson, K., Waldén, M., Bengtsson, H., Ekstrand, J., Häggglund, M., . . . Andersen, T. E. (2016). Role of illness in male professional football: not a

major contributor to time loss. *British Journal of Sports Medicine*, 50(11), 699-702. doi:10.1136/bjsports-2015-095921

Blanch, P., & Gabbett, T. J. (2016). Has the athlete trained enough to return to play safely? The acute:chronic workload ratio permits clinicians to quantify a player's risk of subsequent injury. *British Journal of Sports Medicine*, 50(8), 471-475. doi:10.1136/bjsports-2015-095445

Bowen, L., Gross, A. S., Gimpel, M., Bruce-Low, S., & Li, F.-X. (2019). Spikes in acute:chronic workload ratio (ACWR) associated with a 5–7 times greater injury rate in English Premier League football players: a comprehensive 3-year study. *British Journal of Sports Medicine*, bjsports-2018-099422. doi:10.1136/bjsports-2018-099422

Bowen, L., Gross, A. S., Gimpel, M., & Li, F.-X. (2017). Accumulated workloads and the acute:chronic workload ratio relate to injury risk in elite youth football players. *British Journal of Sports Medicine*, 51(5), 452-459. doi:10.1136/bjsports-2015-095820

Buchheit, M. (2017). Applying the acute:chronic workload ratio in elite football: worth the effort? *British Journal of Sports Medicine*, 51(18), 1325-1327. doi:10.1136/bjsports-2016-097017

Burgess, D. J. (2017). The Research Doesn't Always Apply: Practical Solutions to Evidence-Based Training-Load Monitoring in Elite Team Sports. *I2(s2)*, S2-136. doi:10.1123/ijsp.2016-0608

Cardinale, M., & Varley, M. C. (2017). Wearable Training-Monitoring Technology: Applications, Challenges, and Opportunities. *I2(s2)*, S2-55. doi:10.1123/ijsp.2016-0423

- Carey, D. L., Ong, K., Whiteley, R., Crossley, K. M., Crow, J., & Morris, M. E. (2018). Predictive Modelling of Training Loads and Injury in Australian Football. *17*(1), 49. doi:<https://doi.org/10.2478/ijcss-2018-0002>
- Castellano, J., Casamichana, D., Calleja-González, J., Román, J. S., & Ostojic, S. M. (2011). Reliability and Accuracy of 10 Hz GPS Devices for Short-Distance Exercise. *J Sports Sci Med*, *10*(1), 233-234.
- Clarsen, B. (2014). *Overuse injuries in sport development, validation and application of a new surveillance*. (PhD), Norwegian School of Sport Sciences,
- Clarsen, B., Myklebust, G., & Bahr, R. (2013). Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) Overuse Injury Questionnaire. *British Journal of Sports Medicine*, *47*(8), 495-502. doi:10.1136/bjsports-2012-091524
- Claudino, J. G., Capanema, D. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Med Open*, *5*(1), 28. doi:10.1186/s40798-019-0202-3
- Colby, M. J., Dawson, B., Heasman, J., Rogalski, B., & Gabbett, T. J. (2014). Accelerometer and GPS-Derived Running Loads and Injury Risk in Elite Australian Footballers. *The Journal of Strength & Conditioning Research*, *28*(8), 2244-2252. doi:10.1519/jsc.0000000000000362
- Dalen-Lorentsen, T., Thor Einar Andersen, John Bjørneboe, Markus Vagle, Kevin Nordanger Martin, Michael Kleppen, Morten Wang Fagerland, Benjamin Clarsen. (2019). A cherry tree ripe for picking: The relationship between the acute:chronic workload ratio and health problems. *British Journal of Sports Medicine*.

- Dobbin, K. K., & Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC medical genomics*, 4, 31-31.
doi:10.1186/1755-8794-4-31
- Doll, W. E., & Trueit, D. (2010). Complexity and the health care professions. *Journal of Evaluation in Clinical Practice*, 16(4), 841-848. doi:10.1111/j.1365-2753.2010.01497.x
- Eckard, T. G., Padua, D. A., Hearn, D. W., Pexa, B. S., & Frank, B. S. (2018). The Relationship Between Training Load and Injury in Athletes: A Systematic Review. *Sports Medicine*, 48(8), 1929-1961. doi:10.1007/s40279-018-0951-z
- Ehrmann, F. E., Duncan, C. S., Sindhusake, D., Franzsen, W. N., & Greene, D. A. (2016). GPS and Injury Prevention in Professional Soccer. *The Journal of Strength & Conditioning Research*, 30(2), 360-367.
doi:10.1519/jsc.0000000000001093
- Ekstrand, J., Hägglund, M., & Waldén, M. (2011a). Epidemiology of Muscle Injuries in Professional Football (Soccer). *The American Journal of Sports Medicine*, 39(6), 1226-1232. doi:10.1177/0363546510395879
- Ekstrand, J., Hägglund, M., & Waldén, M. (2011b). Injury incidence and injury patterns in professional football: the UEFA injury study. *British Journal of Sports Medicine*, 45(7), 553-558. doi:10.1136/bjism.2009.060582
- Ekstrand, J., Waldén, M., & Hägglund, M. (2004). A congested football calendar and the wellbeing of players: correlation between match exposure of European footballers before the World Cup 2002 and their injuries and performances during that World Cup. *British Journal of Sports Medicine*, 38(4), 493-497.
doi:10.1136/bjism.2003.009134
- Engebretsen, A. H., Myklebust, G., Holme, I., Engebretsen, L., & Bahr, R. (2010). Intrinsic Risk Factors for Groin Injuries among Male Soccer Players:A

Prospective Cohort Study. *The American Journal of Sports Medicine*, 38(10), 2051-2057. doi:10.1177/0363546510375544

Farmer, J. D., Gallegati, M., Hommes, C., Kirman, A., Ormerod, P., Cincotti, S., . . . Helbing, D. (2012). A complex systems approach to constructing better models for managing financial markets and the economy. *The European Physical Journal Special Topics*, 214(1), 295-324. doi:10.1140/epjst/e2012-01696-9

Faude, O., Rößler, R., & Junge, A. (2013). Football Injuries in Children and Adolescent Players: Are There Clues for Prevention? *Sports Medicine*, 43(9), 819-837. doi:10.1007/s40279-013-0061-x

Fernández Cuevas, I., Carmona, P., Quintana, M., Salces, J., Arnaiz-Lastras, J., & Barrón, A. (2010). *Economic costs estimation of soccer injuries in first and second spanish division professional teams.*

Finch, C. F. (1997). An Overview of Some Definitional Issues for Sports Injury Surveillance. *Sports Medicine*, 24(3), 157-163. doi:10.2165/00007256-199724030-00002

Foster, C., A. Florhaug, J., Franklin, J., Gottschall, L., A. Hrovatin, L., Parker, S., . . . Dodge, C. (2001). *A New Approach to Monitoring Exercise Training* (Vol. 15).

Fousekis, K., Tsepis, E., Poulmedis, P., Athanasopoulos, S., & Vagenas, G. (2011). Intrinsic risk factors of non-contact quadriceps and hamstring strains in soccer: a prospective study of 100 professional players. *British Journal of Sports Medicine*, 45(9), 709-714. doi:10.1136/bjism.2010.077560

Frankish Keith., R. W. (2017). *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press; 2017:

- Freckleton, G., & Pizzari, T. (2013). Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. *British Journal of Sports Medicine*, 47(6), 351-358. doi:10.1136/bjsports-2011-090664
- Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., . . . Meeuwisse, W. H. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *British Journal of Sports Medicine*, 40(3), 193-201. doi:10.1136/bjism.2005.025270
- Gabbett, T., & Domrow, N. (2007). Relationships between training load, injury, and fitness in sub-elite collision sport athletes. *Journal of sports sciences*, 25, 1507-1519. doi:10.1080/02640410701215066
- Gabbett, T. J. (2016). The training—injury prevention paradox: should athletes be training smarter and harder? *British Journal of Sports Medicine*, 50(5), 273-280. doi:10.1136/bjsports-2015-095788
- Gabbett, T. J. (2018). Debunking the myths about training load, injury and performance: empirical evidence, hot topics and recommendations for practitioners. *British Journal of Sports Medicine*, bjsports-2018-099784. doi:10.1136/bjsports-2018-099784
- Gabbett, T. J., Hulin, B. T., Blanch, P., & Whiteley, R. (2016). High training workloads alone do not cause sports injuries: how you get there is the real issue. *British Journal of Sports Medicine*, 50(8), 444-445. doi:10.1136/bjsports-2015-095567
- Gabbett, T. J., & Jenkins, D. G. (2011). Relationship between training load and injury in professional rugby league players. *Journal of Science and Medicine in Sport*, 14(3), 204-209. doi:<https://doi.org/10.1016/j.jsams.2010.12.002>
- Gallo, T., Cormack, S., Gabbett, T., & Lorenzen, C. (2015). Pre-training perceived wellness impacts training output in Australian football players. *Journal of sports sciences*, 34, 1-7. doi:10.1080/02640414.2015.1119295

- Gareth James., D. W., Trevor Hastie., Robert Tibshirani. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company.
- Gonçalves, C. E., Coelho e Silva, M. J., Carvalho, H. M., & Gonçalves, A. (2011). Why do they engage in such hard programs? The search for excellence in youth basketball. *J Sports Sci Med*, *10*(3), 458-464.
- Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K. (2017). Session-RPE Method for Training Load Monitoring: Validity, Ecological Usefulness, and Influencing Factors. *Frontiers in Neuroscience*, *11*(612).
doi:10.3389/fnins.2017.00612
- Harrison, P. W., & Johnston, R. D. (2017). Relationship Between Training Load, Fitness, and Injury Over an Australian Rules Football Preseason. *The Journal of Strength & Conditioning Research*, *31*(10), 2686-2693.
doi:10.1519/jsc.0000000000001829
- Hawkins, R. D., & Fuller, C. W. (1999). A prospective epidemiological study of injuries in four English professional football clubs. *British Journal of Sports Medicine*, *33*(3), 196-203. doi:10.1136/bjism.33.3.196
- Henderson, G., Barnes, C. A., & Portas, M. D. (2010). Factors associated with increased propensity for hamstring injury in English Premier League soccer players. *Journal of Science and Medicine in Sport*, *13*(4), 397-402.
doi:10.1016/j.jsams.2009.08.003
- Higgins, J. P. (2002). Nonlinear systems in medicine. *The Yale journal of biology and medicine*, *75*(5-6), 247-260.
- Hodgson, L., Gissane, C., Gabbett, T. J., & King, D. A. (2007). For Debate: Consensus Injury Definitions in Team Sports Should Focus on Encompassing all Injuries. *Clinical Journal of Sport Medicine*, *17*(3), 188-191.
doi:10.1097/JSM.0b013e3180547513

- Hulin, B. (2017). The never-ending search for the perfect acute: Chronic workload ratio: What role injury definition? *British Journal of Sports Medicine*, *51*, bjsports-2016. doi:10.1136/bjsports-2016-097279
- Hulin, B. T., Gabbett, T. J., Blanch, P., Chapman, P., Bailey, D., & Orchard, J. W. (2014). Spikes in acute workload are associated with increased injury risk in elite cricket fast bowlers. *British Journal of Sports Medicine*, *48*(8), 708-712. doi:10.1136/bjsports-2013-092524
- Hägglund, M., Waldén, M., & Ekstrand, J. (2006). Previous injury as a risk factor for injury in elite football: a prospective study over two consecutive seasons. *British Journal of Sports Medicine*, *40*(9), 767-772. doi:10.1136/bjism.2006.026609
- Hägglund, M., Waldén, M., & Ekstrand, J. (2009). UEFA injury study—an injury audit of European Championships 2006 to 2008. *British Journal of Sports Medicine*, *43*(7), 483-489. doi:10.1136/bjism.2008.056937
- Hägglund, M., Waldén, M., & Ekstrand, J. (2013a). Risk Factors for Lower Extremity Muscle Injury in Professional Soccer: The UEFA Injury Study. *The American Journal of Sports Medicine*, *41*(2), 327-335. doi:10.1177/0363546512470634
- Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., & Ekstrand, J. (2013b). Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *British Journal of Sports Medicine*, *47*(12), 738-742. doi:10.1136/bjsports-2013-092215
- Impellizzeri, F., Rampinini, E., Coutts, A., Sassi, A., & Marcora, S. (2004). Use of RPE-Based Training Load in Soccer. *Medicine & Science in Sports & Exercise*. (Model.IssnPrint). doi:10.1249/01.MSS.0000128199.23901.2
- Jaspers, A., Beéck, T. O. D., Brink, M. S., Frencken, W. G. P., Staes, F., Davis, J. J., & Helsen, W. F. (2018). Relationships Between the External and Internal Training

Load in Professional Soccer: What Can We Learn From Machine Learning?,
13(5), 625. doi:10.1123/ijsp.2017-0299

Johnston, R. J., Watsford, M. L., Kelly, S. J., Pine, M. J., & Spurrs, R. W. (2014).
Validity and Interunit Reliability of 10 Hz and 15 Hz GPS Units for Assessing
Athlete Movement Demands. *The Journal of Strength & Conditioning Research*,
28(6), 1649-1655. doi:10.1519/jsc.0000000000000323

Joseph, A., Fenton, N., & Neil, M. (2006). Predicting football results using Bayesian
nets and other machine learning techniques. *Knowl.-Based Syst.*, 19, 544-553.
doi:10.1016/j.knosys.2006.04.011

Kirkendall, D., & Dvorak, J. (2010). Effective Injury Prevention in Soccer. *The
Physician and sportsmedicine*, 38(1).
doi:<https://doi.org/10.3810/psm.2010.04.1772>

Kotsiantis, S. B. (2007). Supervised machine learning: a review of classification
techniques. *Informatica*(31,249–268).

Laake, P., Olsen, B. R., & Benestad, H. B. (2008). *Forskning i medisin og biofag* (Vol.
2): Gyldendal Akademisk

Lampham, A. B., RM. (1995). The use of artificial intelligence in the analysis of sports
performance: a review of applications in human gait analysis and future
directions for sports biomechanics. *Journal of sports sciences*.
doi:10.1080/02640419508732232

Lathlean, T. J. H., Gatin, P. B., Newstead, S. V., & Finch, C. F. (2019). Absolute and
Relative Load and Injury in Elite Junior Australian Football Players Over 1
Season. 1. doi:10.1123/ijsp.2019-0100

- Liu, Y., Chen, P.-H. C., Krause, J., & Peng, L. (2019). How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature. *Jama*, 322(18), 1806-1816. doi:10.1001/jama.2019.16489
- Lolli, L., Batterham, A. M., Hawkins, R., Kelly, D. M., Strudwick, A. J., Thorpe, R. T., . . . Atkinson, G. (2019). The acute-to-chronic workload ratio: an inaccurate scaling index for an unnecessary normalisation process? *British Journal of Sports Medicine*, 53(24), 1510-1512. doi:10.1136/bjsports-2017-098884
- Malone, J. J., Lovell, R., Varley, M. C., & Coutts, A. J. (2017a). Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *12(s2)*, S2-18. doi:10.1123/ijspp.2016-0236
- Malone, S., Owen, A., Newton, M., Mendes, B., Collins, K. D., & Gabbett, T. J. (2017b). The acute:chronic workload ratio in relation to injury risk in professional soccer. *Journal of Science and Medicine in Sport*, 20(6), 561-565. doi:<https://doi.org/10.1016/j.jsams.2016.10.014>
- Manuel Stein, H. J., Daniel Seebacher, Alexander Jäger, Manuel Nagel, Jürgen Hölsch, Sven Kosub, Tobias Schreck, Daniel A. Keim and Michael Grossniklaus. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *Data open access journal*, 2(1). doi:<https://doi.org/10.3390/data2010002>
- Matthew J. Cross, S. W., Grant Trewartha, Simon P.T. Kemp and Keith A. Stokes. (2016). The Influence of In-Season Training Loads on Injury Risk in Professional Rugby Union. *International Journal of Sports Physiology and Performance*, 11(3). doi:<https://doi.org/10.1123/ijspp.2015-0187>
- Max Kuhn, K. J. (2013). *Applied Predictive Modeling*. Springer New York.
- McCall, A., Dupont, G., & Ekstrand, J. (2018). Internal workload and non-contact injury: a one-season study of five teams from the UEFA Elite Club Injury Study.

British Journal of Sports Medicine, 52(23), 1517-1522. doi:10.1136/bjsports-2017-098473

McCall, A., Fanchini, M., & Coutts, A. (2017). Prediction: The Modern Day Sports Science/Medicine ‘Quest for the Holy Grail’. *International Journal of Sports Physiology and Performance*, 12, 1-11. doi:10.1123/ijsp.2017-0137

Meeuwisse, W. H. (1994). Assessing Causation in Sport Injury: A Multifactorial Model. *Clinical Journal of Sport Medicine*, 4(3), 166-170.

Meeuwisse, W. H., Tyreman, H., Hagel, B., & Emery, C. (2007). A Dynamic Model of Etiology in Sport Injury: The Recursive Nature of Risk and Causation. *Clinical Journal of Sport Medicine*, 17(3), 215-219.
doi:10.1097/JSM.0b013e3180592a48

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28(1), 92-122.
doi:10.1007/s10618-012-0295-5

Mendiguchia, J., Alentorn-Geli, E., & Brughelli, M. (2012). Hamstring strain injuries: are we heading in the right direction? *British Journal of Sports Medicine*, 46(2), 81-85. doi:10.1136/bjism.2010.081695

Michael, R. (uå). An Introduction to Machine Learning for Clinicians. *Academic Medicine Journal of the Association of American Medical Colleges*.
doi:10.1097/ACM.0000000000002792

Miller, T. W. (2015). Sports Analytics and Data Science Winning the Game with Methods and Models.

- Murphy, D. F., Connolly, D. A. J., & Beynon, B. D. (2003). Risk factors for lower extremity injury: a review of the literature. *British Journal of Sports Medicine*, 37(1), 13-29. doi:10.1136/bjism.37.1.13
- Nielsen, R. Ø., Malisoux, L., Møller, M., Theisen, D., & Parner, E. T. (2016). Shedding Light on the Etiology of Sports Injuries: A Look Behind the Scenes of Time-to-Event Analyses. *Journal of Orthopaedic & Sports Physical Therapy*, 46(4), 300-311. doi:10.2519/jospt.2016.6510
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-1379. doi:[https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
- Petersen, J., Thorborg, K., Nielsen, M. B., Budtz-Jørgensen, E., & Hölmich, P. (2011). Preventive Effect of Eccentric Training on Acute Hamstring Injuries in Men's Soccer: A Cluster-Randomized Controlled Trial. *The American Journal of Sports Medicine*, 39(11), 2296-2303. doi:10.1177/0363546511419277
- Pfiffmann, D., Herbst, M., Ingelfinger, P., Simon, P., & Tug, S. (2016). Analysis of Injury Incidences in Male Professional Adult and Elite Youth Soccer Players: A Systematic Review. *Journal of athletic training*, 51(5), 410-424. doi:10.4085/1062-6050-51.6.03
- Polglaze, T., Dawson, B., Hiscock, D. J., & Peeling, P. (2015). A Comparative Analysis of Accelerometer and Time-Motion Data in Elite Men's Hockey Training and Competition. *10*(4), 446. doi:10.1123/ijssp.2014-0233
- Quatman, C. E., Quatman, C. C., & Hewett, T. E. (2009). Prediction and prevention of musculoskeletal injury: a paradigm shift in methodology. *British Journal of Sports Medicine*, 43(14), 1100-1107. doi:10.1136/bjism.2009.065482

- Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krstrup, P., & Rebelo, A. (2019). Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: a systematic review. *Research in Sports Medicine*, 1-22. doi:10.1080/15438627.2019.1686703
- Rampinini, E., Alberti, G., Fiorenza, M., Riggio, M., Sassi, R., Borges, T. O., & Coutts, A. J. (2015). Accuracy of GPS Devices for Measuring High-intensity Running in Field-based Team Sports. *Int J Sports Med*, 36(01), 49-53. doi:10.1055/s-0034-1385866
- Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health*, 61(11), 933-937. doi:10.1136/jech.2006.054254
- Rico-González, M., Los Arcos, A., Nakamura, F., Moura, F., & Pino Ortega, J. (2019). The use of technology and sampling frequency to measure variables of tactical positioning in team sports: a systematic review. *Research in Sports Medicine*. doi:10.1080/15438627.2019.1660879
- Robert P. Schumaker., O. K. S., Hsinchun Chen. (2010). *Sports Data Mining* (1 ed.).
- Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., . . . D'Hondt, E. (2020). A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine & Science in Sports & Exercise*, 1. doi:10.1249/MSS.0000000000002305
- Ruddy, J. D., Cormack, S. J., Whiteley, R., Williams, M. D., Timmins, R. G., & Opar, D. A. (2019). Modeling the Risk of Team Sport Injuries: A Narrative Review of Different Statistical Approaches. *Front Physiol*, 10, 829. doi:10.3389/fphys.2019.00829

- Sampson, J. A., Fullagar, H. H. K., & Murray, A. (2017). Evidence is needed to determine if there is a better way to determine the acute:chronic workload. *British Journal of Sports Medicine*, 51(7), 621-622. doi:10.1136/bjsports-2016-097085
- Schwellnus, M., Soligard, T., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, H. P., . . . Engebretsen, L. (2016). How much is too much? (Part 2) International Olympic Committee consensus statement on load in sport and risk of illness. *British Journal of Sports Medicine*, 50(17), 1043-1052. doi:10.1136/bjsports-2016-096572
- Scott, M. T. U., Scott, T. J., & Kelly, V. G. (2016). The Validity and Reliability of Global Positioning Systems in Team Sport: A Brief Review. *The Journal of Strength & Conditioning Research*, 30(5), 1470-1490. doi:10.1519/jsc.0000000000001221
- Scott, T. J., Black, C. R., Quinn, J., & Coutts, A. J. (2013). Validity and Reliability of the Session-RPE Method for Quantifying Training in Australian Football: A Comparison of the CR10 and CR100 Scales. *The Journal of Strength & Conditioning Research*, 27(1), 270-276. doi:10.1519/JSC.0b013e3182541d2e
- Shmueli, G. (2010). To Explain or to Predict? *Statist. Sci.*, 25(3), 289-310. doi:10.1214/10-STS330
- Smoliga, J. M., & Zavorsky, G. S. (2017). Team Logo Predicts Concussion Risk: Lessons in Protecting a Vulnerable Sports Community from Misconceived, but Highly Publicized Epidemiologic Research. *Epidemiology*, 28(5), 753-757. doi:10.1097/ede.0000000000000694
- Soligard, T., Schwellnus, M., Alonso, J.-M., Bahr, R., Clarsen, B., Dijkstra, H. P., . . . Engebretsen, L. (2016). How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *British*

Journal of Sports Medicine, 50(17), 1030-1041. doi:10.1136/bjsports-2016-096581

Soto-Valero, C. (2016). Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport*, 15, 91-112. doi:10.1515/ijcss-2016-0007

Stern, B. D., Hegedus, E. J., & Lai, Y.-C. (2020). Injury prediction as a non-linear system. *Physical Therapy in Sport*, 41, 43-48.
doi:<https://doi.org/10.1016/j.ptsp.2019.10.010>

Sverre Mæhlum, R. B. (2017). Idrettskade i store medisinske leksikon. Hentet 11.05.2020 fra <https://sml.snl.no/idrettsskade>

Theron, G. (2020). *The use of Data Mining for Predicting Injuries in Professional Football Players* (Master Thesis), University of Oslo,

van Dyk, N., & Clarsen, B. (2017). Prevention forecast: cloudy with a chance of injury. *British Journal of Sports Medicine*, 51(23), 1646-1647. doi:10.1136/bjsports-2017-097861

van Dyk, N., van der Made, A. D., Timmins, R. G., Opar, D. A., & Tol, J. L. (2018). There is strength in numbers for muscle injuries: it is time to establish an international collaborative registry. *British Journal of Sports Medicine*, 52(19), 1228-1229. doi:10.1136/bjsports-2016-097318

van Mechelen, W. (1997). Sports Injury Surveillance Systems. *Sports Medicine*, 24(3), 164-168. doi:10.2165/00007256-199724030-00003

van Mechelen, W., Hlobil, H., & Kemper, H. C. G. (1992). Incidence, Severity, Aetiology and Prevention of Sports Injuries. *Sports Medicine*, 14(2), 82-99.
doi:10.2165/00007256-199214020-00002

- Verghese, A., Shah, N. H., & Harrington, R. A. (2018). What This Computer Needs Is a Physician: Humanism and Artificial Intelligence. *Jama*, *319*(1), 19-20. doi:10.1001/jama.2017.19198
- Waldén, M., Hägglund, M., & Ekstrand, J. (2005a). Injuries in Swedish elite football—a prospective study on injury definitions, risk for injury and injury pattern during 2001. *Scandinavian Journal of Medicine & Science in Sports*, *15*(2), 118-125. doi:10.1111/j.1600-0838.2004.00393.x
- Waldén, M., Hägglund, M., & Ekstrand, J. (2005b). UEFA Champions League study: a prospective study of injuries in professional football during the 2001-2002 season. *Br J Sports Med*, *39*(8), 542-546. doi:10.1136/bjsm.2004.014571
- Weston, M. (2018). Training load monitoring in elite English soccer: a comparison of practices and perceptions between coaches and practitioners. *Science and Medicine in Football*, *2*(3), 216-224. doi:10.1080/24733938.2018.1427883
- Windt, J., & Gabbett, T. J. (2017). How do training and competition workloads relate to injury? The workload—injury aetiology model. *British Journal of Sports Medicine*, *51*(5), 428-435. doi:10.1136/bjsports-2016-096040

Tabelloversikt

Tabell 1: Oversikt over hvilke variabler som ble hentet fra GPS dataene, hva de måler og hvilken måleenhet som benyttes.....	31
Tabell 2: Tabellen viser skadedata for begge skadedefinisjonene og totalt antall dager spillerne hadde fravær fra skade gjennom sesongen.	40
Tabell 3: Tabellen viser antall registrerte skader fordelt på lokalisasjon og antallet akutte belastningsskader og belastningsskader.....	41
Tabell 4: Tabellen viser de fire forskjellige maskinlæringsmodellenes evne til å predikere skade for «ikke-kontakt skade» og «ikke-kontakt skade fravær», presenter med presisjon, sensitivitet, F1 og AUK score. Forkortelser: AUK= Areal under kurven.	43
Tabell 5: Tabellen viser gjennomsnitt og standardavvik for distanse, playerload™, acceleration load, V4 distanse, V5 distanse og høyhastighetsløp. (Gjennomsnitt er beregnet fra hele sesongen).	46

Figuroversikt

Figur 1: Stegene for utviklingen av skadeforebyggende tiltak (van Mechelen, 1992) .	11
Figur 2: Meeuwisse og hans medarbeidere 1994, første multifaktorelle skademodell.	15
Figur 3: Bahr og Krosshaugs etiologimodell (Bahr & Krosshaug, 2005).	15
Figur 4: Meeuwisse og hans medarbeidere (2007) sin dynamiske etiologimodell.	16
Figur 5: Belastnings-etnologiskademodellen (Windt & Gabbett, 2017).	17
Figur 6: Kompleks modell for idrettsskader. Gruppen med sirkler nederst skal inneholde ulike variabler, disse utgjør determinantene. Variablene med de mørke sirkler har flere interaksjoner enn variablene med lysere sirkler og har en større påvirkning på utfallet. De med stiplede linje representerer svakere interaksjoner, mens de med mørk linje indikerer sterkere interaksjon mellom variablene (Bittencourt et al., 2016).	24
Figur 7: Eksempel på en GPS-enhet med innebygd akselerometer og på en spesialdesignet vest som tilsvarer den spillerne brukte.	31
Figur 8: Flytdiagram som viser antall spillere, treningsdata og antall skader som ble inkludert i prosjektet. Forkortelser: Ikke-kontakt F= Ikke kontakt skade som fører til fravær.	39
Figur 9: Figuren viser forholdet mellom totalt antall økter, og økter med registrert skade.	39
Figur 10: Oversikt over tidspunkt for skadene og skadenes varighet gjennom 2019 sesongen. Søylen for IK-F skadene viser både varigheten for spillerens skade, samt antall dager spilleren hadde fravær fra trening eller kamp. IK-skadene er markert som varigheten på skaden, og ikke antall økter med fravær. Grunnen er at IK-skade per definisjon ikke forårsaker fravær fra trening eller kamp. Forkortelser: IK-skade= Ikke kontakt skade. IK-F = Ikke kontakt skade som fører til fravær.	42
Figur 11: ROC kurve som viser arealet under kurven på 0.5 for maskinlæringsmodellen. Alle modellen fikk lik AUK. X-aksen representerer tilfeller der modellen har predikert riktig. Y-aksen viser tilfeller hvor modellen har predikert feil. AUK= Areal under kurven. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest	44
Figur 12: Tettheten til AUK for IK-skade ved 1000 gjennomkjøringer av algoritmen. Forkortelser: AUK= arealet under kurven. IK= Ikke kontakt skader. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest	45
Figur 13: Tettheten til AUK for IK-F skade fravær ved 1000 gjennomkjøringer av modellene. Forkortelser: AUK= Arealet under kurven. IK-F = Ikke kontakt fravær. SVM = Support vector machine. DT = Decision tree. LR = Logistisk regression. RF = Random forest	45

Vedlegg

- A** Skjema for søknad om godkjenning av forskningsprosjekt
- B** Prosjektbeskrivelse: Machine learning in professional football research
- C** Samtykkeskjema
- D** NSD Norsk senter for forskningsdata

Vedlegg A

Skjema for søknad om godkjenning av forskningsprosjekt Etisk komite for idrettsvitenskapelig forskning på mennesker – NIH 22102017

Retningslinjer for søknad til Etisk komite for idrettsvitenskapelig forskning på mennesker ved Norges idrettshøgskole må leses før utfylling av skjemaet. Søknadsskjemaet og vedlegg (se pkt 6) skal være pdf-dokumenter som sendes samlet som ett pdf-dokument innen angitt tidsfrist. Vedleggene skal nummereres i henhold til pkt 6 i skjemaet.

1. Generelle opplysninger

1.1 Prosjektleder

Prosjektleder har ansvaret for den daglige driften av forskningsprosjektet og skal ha nødvendige forskningskvalifikasjoner (hovedregel dr. grad eller tilsvarende) og erfaring i forskningsetikk, herunder personvern og informasjonssikkerhet. Prosjektleder skal informere seksjonsleder om forskningsprosjektet, herunder om søknad til etisk komite ved NIH.

Navn: Thor Einar Andersen

Stilling: Professor

Seksjon: Seksjon for
idrettsmedisinske fag

1.2 Forskningsansvarlig

Seksjonsleder skal oppføres som forskningsansvarlig og skal være informert om søknad til NIHs etiske komite

Navn: Sigmund Andersen

1.3 Prosjekttittel

Norsk tittel

Kortfattet, allment forståelig

Machine learning i profesjonell fotball

Vitenskapelig tittel

Gjerne engelsk, maksimalt 1000 tegn

Machine learning and injuries in professional football

1.4 Initiativtaker

Hvem er initiativtaker til prosjektet?

- X Prosjektleder eller andre med ansettelse med ved NIH
- Ekstern forsker/forskningsinstitusjon
- Myndighet, firma (Oppdragsforskning)

1.5 Utdanningsprosjekt

Er prosjektet del av en master eller doktorgrad?

- X Ja
- Nei

1.6 Prosjektmedarbeidere

Prosjektmedarbeidere er personer som bidrar med selvstendig vitenskapelig arbeid i et forskningsprosjekt

Navn	Stilling	Institusjon	Akademisk grad	Prosjektrolle
Torstein Dalen-Lorentsen	Stipendiat	NIH	MSc	Daglig leder
Thor Einar Andersen	Professor	NIH	MD, PhD	Prosjektleder
Anders Larsen	Masterstudent	NIH	BSc	Student
Vera Hermine Goebel	Professor	UiO / IFI / DMMS	PhD	Prosjektmedarbeider
Thomas Peter Plagemann	Professor	UiO / IFI / DMMS	PhD	Prosjektmedarbeider
Garth Theron	Masterstudent	UiO / IFI / DMMS	BSc	Student

1.7 Tidsramme for prosjektet

Prosjektstart er tidspunkt for inkludering av forskningsdeltakere og innsamling av data. Prosjektsslutt er tidspunktet tillatelsen til å behandle personopplysninger går ut i henhold til NSD godkjenning. Styret ved NIH har vedtatt at forskningsdata skal lagres i 5 år etter prosjektsslutt for etterprøvnbarhet og kontroll. Dette innebærer at du må angi en prosjektperiode som varer 5 år etter at prosjektet er avsluttet og at NSD har fått denne opplysningen i søknad og godkjent dette. Godkjenning fra NSD trenger ikke å foreligge, men søknad skal være sendt og kopi av NSD-søknad vedlegges.

Prosjektstart dato: 01.09.19

Prosjektsslutt dato: 31.12.2023.

1.8 Samarbeid med utlandet

Har prosjektet noen form for samarbeid med utlandet?

- Ja
- X Nei

1.9 Annet prosjekt med betydning for vurderingen

Er det noe annet prosjekt som kan ha betydning for vurderingen av det aktuelle prosjektet? F.eks. et hovedprosjekt eller delprosjekt

- Ja, angi tittel på prosjekt og hvor prosjektet er forankret (prosjektleder/institusjon)
- X Nei

2. Prosjektopplysninger

2.1 Oppsummering av forskningsprosjektet

Kort prosjektbeskrivelse

Hvilken ny kunnskap skal forskningen gi? Hvilken forskningsdesign og –metode skal brukes? Gi en allment forståelig og kortfattet beskrivelse av hvilke forskningsspørsmål prosjektet skal besvare og hvordan de skal besvares. Formålet med prosjektet må komme klart frem.

Formålet med prosjektet er å innhente ny kunnskap om hvordan skader oppstår, og hvilke faktorer som påvirker sannsynligheten for skader i profesjonell fotball. Det har vært mye forskning på området, men kun ved bruk av enkeltfaktorer som risikofaktor. Det er bred enighet at skader oppstår som et resultat av en rekke faktorer og forhold som til sammen fører til skade. Forskningen på enkeltfaktorer baserer seg på en reduksjonistisk tankegang som ikke er forenlig med skaders multifactorielle natur og er derfor ikke optimal for å få forståelse for hvordan skader oppstår og hvordan vi kan forhindre dem. På bakgrunn av dette ønsker vi å undersøke bruk av muligheter multidimensjonal statistikk og maskinlæring for å se hvilke faktorer som er forbundet med skader i profesjonell fotball.

2.2 Forskningsdata

Sensitive personopplysninger? (Rasemessig eller etnisk bakgrunn, politisk, filosofisk eller religiøs oppfatninger, person mistenkt, siktet, tiltalt eller dømt for en straffbar handling, helseforhold, seksuelle forhold eller medlemskap fagforeninger)

- Ja
 Nei

Tidligere registrerte personopplysninger?

- Ja
 Nei

Nye personopplysninger

Personopplysninger som skal samles inn direkte fra studiepopulasjonen, ved f.eks. klinisk undersøkelse, intervensjon eller spørreskjema.

- Ja
 Nei

Humant biologisk materiale

Materiale som allerede er samlet inn eller som skal samles inn i prosjektet. Humant biologisk materiale er organer, deler av organer, celler og vev og bestanddeler av slikt materiale fra levende og døde mennesker.

- Ja
 Nei

2.3 Studiepopulasjonen

Antall forskningsdeltakere og styrkeberegning

Oppgi antall forskningsdeltakere i Norge og evt. I utlandet. Begrunn antallet/eventuelt valg av kjønn. Redegjør for styrkeberegning ved statistiske analysemetoder.

Vi ønsker å inkludere samtlige spillere med fast kontrakt i vår samarbeidsklubb.

Beskrivelse av forskningsdeltakere/utvalg

Kryss av og beskriv hvorfor disse personene skal inkluderes

- Personer mellom 16 og 18 år
- X Personer over 18 år
- Personer med redusert samtykkekompetanse
- Mindreårige under 16 år
- Andre personer i en sårbar eller avhengig situasjon

F.eks. innsatte i fengsel, soldater, ansatte, elever (Det kreves spesiell begrunnelse for å inkludere personer i en sårbar eller avhengig situasjon, fordi det for disse kan være vanskelig å ivareta prinsippet om frivillig deltakelse)

Beskrive under hvorfor disse personene skal inkluderes

Vi inkluderer alle i målgruppen.

2.4 Forskningsmetode

Metode for analysering av data

- X Statistiske (kvantitative) analysemetoder
- Fortolkende (kvalitative) analysemetoder

Metode for innhenting av data

- X Fysiske tester (eks. opplæringsprogram, treningsprogram)
- X Kliniske undersøkelser
- X Andre intervensjoner over tid (eks. pre- post målinger)
- X Spørreskjema
- Intervju
- Observasjon

2.5 Begrunnelse for valg av data og metode

Redegjør for den faglige og vitenskapelige begrunnelsen for valg av data og metode

I tillegg til å undersøke faktorer relatert til skade er bakgrunnen for studien å utforske hvilken metode av multidimensjonal statistikk som passer vårt datasett på best måte. Denne metoden setter ingen begrensning på mengde data og vi vil inkludere alle parameter samlet inn i klubbregi. Disse data vil i stor grad komme fra mikrosensorer som spillerne benytter i trening og kamp. Disse gir data på blant annet distanse, hastighet, akselerasjon og krefter i forskjellige plan. I tillegg vil vi bruke selvrappertert belastning og resultater på tester fra spillerne. Utfallsmål vil være helsedata registrert av klubbens medisinske personell, samt prestasjonsdata fra mikrosensorer og statistikk fra kamper.

Da dette vil være en eksplorativ studie vil metoden kunne endre seg mens den blir til. Våre samarbeidspartnere på UiO vil være de som primært vurderer hvilken type statistisk metode som vil være mest hensiktsmessig. Dette blir gjort fortløpende mens data blir lagt inn i systemet.

3. Informasjon, samtykke og personvern

3.1 Samtykke vil bli innhentet

Hvis ja må Informasjonsskriv legges ved. Samtykket til deltakelse i forskning skal som hovedregel være informert, frivillig, uttrykkelig og dokumenterbart. Forespørsel om deltakelse og samtykkeerklæring skal utformes i samsvar med mal for informasjonsskriv. Lenker til maler for informasjonsskriv finner du på REK/NSDs sine hjemmesider. Det skal opplyses om at forskningsdata vil bli lagret i 5 år for etterprøvnbarhet og kontroll.

- X Ja
 Nei

3.2 Samtykke er allerede innhentet

Hvis ja må tidligere godkjent informasjonsskriv legges ved.

- X Ja
 Nei

3.3 Det søkes om fritak fra kravet om å innhente samtykke

- Ja
 X Nei

4. Avveining av nytte og risiko ved prosjektet

4.1 Fordeler

Angi fysisk, psykisk, sosial og/eller praktisk fordel/nytte/gagn nå eller i fremtida for den enkelte pasient/deltaker, grupper av personer, samfunnet og/eller vitenskapen.

Studien vil kunne hjelpe klubben forstå hvordan forskjellige parameter vil henge sammen med både skader og prestasjon, og vil dermed kunne potensielt bruke dette som en konkurransefordel i etterkant av studien.

4.2 Ulemper

Angi fysisk, psykisk, sosial og/eller praktisk risiko/skade/ubehag/belastning/uileilighet nå eller i fremtida for den enkelte pasient/deltaker, grupper av personer, samfunnet og/eller miljø

Ingen ulemper blir påført spillerne da vi kun vil bruke data som uansett ville blitt samlet inn i klubbsetting. Data vil bli behandlet aidentifisert og på en forsvarlig måte. Henviser til NSD-godkjenning for videre detaljer.

4.3 Tiltak

Redegjør for tiltak for å ivareta og beskytte deltakerne i forskningsprosjektet og for å begrense mulig risiko/ulempe. Diskuter beredskap ved uventede hendelser og uventede funn der dette er aktuelt. Tiltak for å ivareta og beskytte deltakere i prosjektet kan for eksempel være, styrking av samtykkekompetanse, ekstra beskyttelse av deltakere i en sårbar eller avhengig situasjon, sikring av konfidensialitet ved kvalitative metoder og lite antall deltakere, eksklusjonskriterier, klinisk forundersøkelse, beredskap, interimsanalyser eller oppfølging av deltakere.

Spillerne vil få en grundig innføring i personvern knyttet til studien og tilstrekkelig informasjon rundt sine rettigheter i forbindelse med deltagelse.

4.4 Forsvarlighet

Hvorfor er det forsvarlig å gjennomføre prosjektet? Gi en begrunnet avveining av fordelene og ulempene ved forskningsprosjektet.

Prosjektet er forsvarlig å gjennomføre da selve innsamlingen av data vil ikke være merkbar i spillernes hverdag. Data vil bli behandlet aidentifisert og på en forsvarlig måte. Fordelene for spillerne vil være at klubben vil få nyttig innsikt i hvordan de kan unngå skader på spillerne, og potensielt øke prestasjonen på sikt. Henviser til NSD-godkjenning for videre detaljer.

5. Vurdering av andre instanser og interesser

5.1 Vurdering av andre instanser

Det skal som hovedregel ikke sendes søknad til REK og NIHs etiske komite samtidig. I de tilfeller søknad er sendt til REK, vil etisk komite avvente behandling av søknaden inntil det foreligger et REK vedtak. Er det tvil om prosjektet skal behandles av

REK i henhold til Helseforskningsloven, skal skjemaet for fremleggelsesvurdering sendes inn til REK. REK sitt svar på fremleggelsesvurderingen/vedtak fra REK skal vedlegges søknaden

Er det sendt søknad til REK?

- Ja
 X

Nei

Vurdering av andre instanser skal vedlegges hvis det anses relevant for søknaden.

Prosjektet har blitt vurdert: NSD **Godkjent**

5.2 Interesser

Finansieringskilder

Hvem finansierer prosjektet? Ved oppdragsforskning skal økonomisk avtale vedlegges eller ettersendes NIH SIM/Senter for idrettsskedeforskning

Kompensasjon til forskningsdeltakere

Eventuell kompensasjon for utgifter, tapt arbeidsfortjeneste, tidsbruk, ulempe eller annet

X

Eventuelle interessekonflikter for prosjektleder/-medarbeidere

Det skal redegjøres for eventuelle bindinger til oppdragsgiver, eierinteresser, styreverv, aksjeinteresser

Torstein Dalen-Lorentsen – stipendiat NIH er også ansatt som fysisk trener i den aktuelle klubben.

6. Vedlegg

Hvert vedlegg skal være et pdf-dokument som nummereres som følger:

Vedlegg 1 Forskningsprotokoll/Prosjektplan

Vedlegg 2 Samtykkeskriv

Vedlegg 3 Søknad NSD, Ev godkjenning fra NSD hvis denne foreligger

Vedlegg 4 Spørreskjema hvis aktuelt

Vedlegg 5 CV for prosjektleder hvis ikke ansatt ved NIH

Vedlegg 6 Ev korrespondanse med REK (svar på fremleggesskjema/vedtak om at prosjektet falt utenfor Helseforskningsloven) Kun aktuelt dersom prosjektleder har vært i tvill om/ment at prosjektet faller innenfor Helseforskningsloven.

Vedlegg X Annen dokumentasjon og opplysninger som er nødvendig for å få en full forståelse for søknaden

Søknadsskjema og vedlegg (i pdf-format) skal samles i ett pdf-dokument for innsendelse

7. Ansvarserklæring

Jeg erklærer at prosjektet vil bli gjennomført

- X I henhold til gjeldende lover, forskrifter og retningslinjer
- X I samsvar med opplysninger gitt i denne søknaden
- X I samsvar med eventuelle vilkår for godkjenning gitt av NIHs etiske komite, NSD og ev andre godkjenninginstanser

Vedlegg B

Machine learning in professional football research

Background

Injuries among footballers are prevalent and each player has to expect somewhere around two injuries per season (Ekstrand, Hägglund, & Waldén, 2011). Injuries are detrimental for long term player development (Gabbett, Hulin, Blanch, & Whiteley, 2016), and elicit a negative effect on team performance (Hägglund et al., 2013). Injuries represent also a financial burden for professional clubs and are estimated to cost each English Premier League club an average of 8,85 million GBP per season (JLT group, 2017).

There has been a large amount of work trying to discover the potential risk factors for injuries in football. Research has found several associations with different injury types and locations ((Freckleton & Pizzari, 2013; McCall et al., 2014; Nilstad, Andersen, Bahr, Holme, & Steffen, 2014)). This has, however, solely been analyzed as single parameters as risk factors and has not taken the relationship between parameters or other factors into the analyses. This reductionist approach does not take the multifactorial nature of injuries into account and is possible sub-optimal for getting an understanding of how injuries occur, and how we can prevent them.

Thus, the focus on risk factor research is shifting towards multi-parameter statistics (Stein et al., 2017). Either as a combination of known risk factors, such as training load (TL) increases and scapular dyskinesis in handball players (Møller et al., 2017), or as more complex approaches as machine learning (Rossi et al., 2018)}. The introduction of machine learning and similar approaches allows us to include more parameters (and more metadata) into our analyses.

Advances in player tracking technology give us accurate and valid data on players training load in training and match (Scott, Black, Quinn, & Coutts, 2013). Global Positioning Systems (GPS) and accelerometer devices are typically worn every training and match and provide us positional and physical data sampled at 10-1000Hz. These systems create a large amount of data that practitioners and researchers use for training load management and risk factor research. However, this data is typically analyzed as separate parameters and not in relation to each other or other data (such as age, wellness, strength measurements etc).

Professional clubs are interested in knowing what measures they can employ to reduce injuries and increase performance (Kirkendall, 2010). In order to know this, the picture needs to get clearer on which parameters and which patterns that influences injuries and performance.

The amount of data collected each day in the professional clubs is massive and it has a large potential for investigation. To analyze these data, a traditional reductionist approach is not sufficient, and a more advanced method is called for (Bittencourt et al., 2016).

Methods

This will be a prospective cohort study. The team will train and compete as usual, with no interference from the research group. The data collection will take place during the 2019 season (March to November). We will use a GPS unit with an accelerometer (Catapult OptimEye X4). The GPS unit is placed in a specially designed vest, where the device is placed in a pocket between the shoulder blades

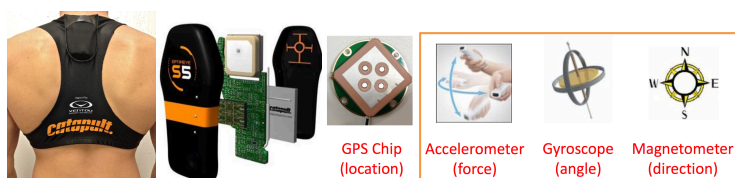


Figure 2. An example of an advanced GPS device, integrated with inertial sensors

Research questions

The aim of the study is to uncover hidden trends and develop new knowledge about the complexity of injuries and performance in football, using data collection results. The aim is to discover patterns the club can use in their training methodology in the future for either injury prevention or performance enhancement.

References

- Ekstrand, J., Hägglund, M., & Waldén, M. (2011). Injury incidence and injury patterns in professional football: the UEFA injury study. *British Journal of Sports Medicine*, *45*(7), 553–558. <http://doi.org/10.1136/bjsm.2009.060582>
- Freckleton, G., & Pizzari, T. (2013). Risk factors for hamstring muscle strain injury in sport: a systematic review and meta-analysis. *British Journal of Sports Medicine*, *47*(6), 351–358. <http://doi.org/10.1136/bjsports-2011-090664>
- Gabbett, H. T., Hulin, B. T., Blanch, P., & Whiteley, R. (2016). High training workloads alone do not cause sports injuries: how you get there is the real issue. *British Journal of Sports Medicine*, *50*(8), 444–445. <http://doi.org/10.1136/bjsports-2015-095567>
- JLT Speciality Group, Injury Report Premier League 2017. (2017). JLT Sports Injury Index, 1–28.
- Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., & Ekstrand, J. (2013). Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. *British Journal of Sports Medicine*, *47*(12), 738–742. <http://doi.org/10.1136/bjsports-2013-092215>
- McCall, A., Carling, C., Nédélec, M., Davison, M., Le Gall, F., Berthoin, S., & Dupont, G. (2014). Risk factors, testing and preventative strategies for non-contact injuries in professional football: current perceptions and practices of 44 teams from various premier leagues. *British Journal of Sports Medicine*, *48*(18), bjsports–2014–093439–1357. <http://doi.org/10.1136/bjsports-2014-093439>
- Møller, M., Nielsen, R. O., Attermann, J., Wedderkopp, N., Lind, M., Sørensen, H., & Myklebust, G. (2017). Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *British Journal of Sports Medicine*, *51*(4), 231–237. <http://doi.org/10.1136/bjsports-2016-096927>
- Nilstad, A., Andersen, T. E., Bahr, R., Holme, I., & Steffen, K. (2014). Risk factors for lower extremity injuries in elite female soccer players. *The American Journal of Sports Medicine*, *42*(4), 940–948. <http://doi.org/10.1177/0363546513518741>
- Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernandez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLoS ONE*, *13*(7), e0201264–15. <http://doi.org/10.1371/journal.pone.0201264>
- Scott, T. J., Black, C. R., Quinn, J., & Coutts, A. J. (2013). Validity and reliability of the session-RPE method for quantifying training in Australian football: a comparison of the CR10 and CR100 scales. *Journal of Strength and Conditioning Research / National Strength & Conditioning Association*, *27*(1), 270–276. <http://doi.org/10.1519/JSC.0b013e3182541d2e>
- Stein, M., Janetzko, H., Seebacher, D., Jäger, A., Nagel, M., Hölsch, J., et al. (2017). How to Make Sense of Team Sport Data: From Acquisition to Data Modeling and Research Aspects. *Data*, *2*(1), 2–23. <http://doi.org/10.3390/data2010002>

Vedlegg C

Do you want to participate in the research project ”Machine learning in professional football”?

This is a query for you to take part in a research project where the aim is to identify patterns leading to performance or injury, using data collected in a professional football club. In this information letter, you will get information on the targets of the project, and what this will mean for you as a potential research participant.

Aim

Injuries among footballers are prevalent and each player has to expect somewhere around two injuries per season. Injuries are detrimental for long term player development, and elicit a negative effect on team performance. Research has found several associations with different injury types and locations. This has, however, solely been analyzed as single parameters as risk factors and has not taken the relationship between parameters or other factors into the analyses. This reductionist approach does not take the multifactorial nature of injuries into account and is sub-optimal for getting an understanding of how injuries occur, and how we can prevent them. Thus, the focus on risk factor research is shifting towards multi-parameter statistics. The introduction of machine learning and similar approaches allows us to include more parameters (and more metadata) into our analyses. This project aims to involve machine learning as a methodological tool to investigate patterns leading to injury.

Who is responsible for the project?

Oslo Sports Trauma Research Center is responsible for the project.

This is a part of a research collaboration between the Oslo Sports Trauma Research Center and the Distributed Multimedia Systems (DMMS) research group at the faculty of Mathematics and Natural Sciences, University of Oslo.

Why are you invited to participate?

Every player in with a professional contract with Strømsgodset TF is invited

What does it mean for you as a participant?

All data will be collected through the clubs own coaching and medical staff as normal. No change in routine or practice. The data used in the project will purely be observational and you as a participant will not be affected in the daily practice. We will collect health problem data through the club's own injury surveillance systems. This includes all health related complaints by the player to the club's medical and performance staff.

Participation is optional

participation in this project is optional. If you choose to participate, you could withdraw at any time, without providing a reason. Withdrawal to the project will not elicit any negative consequence for you at the club or at any other circumstance.

Your privacy – How we store and use your information

We will only use your information in the settings and purposes that is included in this information letter. We will treat the information strictly confidential and in alignment with the General Data Protection Regulation (GDPR).

- *Only Anders Larsen (master student OSTRC) and Torstein Dalen-Lorentsen (Fitness coach Strømsgodset and PhD-candidate OSTRC) have access to the data that is not yet unidentified.*

The identifiable data (Name etc.) will be exchanged with a participant code and handed over to the research group at DMMS (approximately 5-10 researchers will work on the project).

- *The original data will be stored through the clubs own server with the routines of the club.*
- *The results from the study will be in a manner where the participants will not be recognizable.*

What happens to your information when the project is finished?

Project will end on December 31, 2023. All identifiable data will be deleted after this date. Unidentifiable data will be stored on the OSTRC server. Only project leader (Torstein Dalen-Lorentsen) will have access.

Your rights

As long as you are identifiable in the data, your rights are:

- Obtain all information about your self
- To correct all information about your self
- To delete all information about your self
- To receive a copy of all information about your self
- To send a complaint to the Data protection officer at the Norwegian School of Sports Sciences or to The Norwegian Data Protection Authority regarding the usage of your personal information.

What gives us the right to use your personal information?

We use your personal information based on your written consent.

On assignment from the Oslo Sports Trauma Research Center has NSD - Norwegian Centre for Research Data assessed that the usage of personal information is aligned with the GDPR

Where can I obtain more information?

If you have questions regarding the project or wish to use your rights, please contact

- *Torstein Dalen-Lorentsen, PhD-Candidate, Oslo Sports Trauma Reserach Center, torstein.dalen@nih.no, +4793841844. Or Thor Einar Andersen, Professor, Oslo Sports Trauma Research Center, t.e.andersen@nih.no.*
- Our GDPR responsible Karine Justad, Norwegian School of Sport Sciences, +4797536704. Karine.justad@nih.no
- NSD - Norwegian Centre for Research Data, personverntjenester@nsd.no, +4755582117

Best regards
Project Leader
Torstein Dalen-Lorentsen

Declaration of consent

I have received and understood information about the project Machine learning in professional football, and I have had access to ask questions regarding the project information.

I hereby give my consent to:

- Participate in the project (Information is collected through the clubs own routines and practices).

I give consent that my personal information can be used until the project is finished (ca. dec 2023)

(Signed by participant, Date)

Vedlegg D

NSD NORSK SENTER FOR FORSKNINGSDATA**NSD sin vurdering****Prosjekttittel**

Machine learning in professional football

Referansenummer

722773

Registrert

03.04.2019 av Torstein Dalen-Lorentsen - torstein.dalen@nih.no

Behandlingsansvarlig institusjon

Norges idrettshøgskole / Senter for idrettsskedeforskning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Torstein Dalen-Lorentsen, torstein.dalen@nih.no, tlf: 93841844

Type prosjekt

Forskerprosjekt

Prosjektperiode

02.04.2019 - 31.12.2023

Status

13.06.2019 - Vurdert

Vurdering (1)

13.06.2019 - Vurdert

Prosjektleder skal innhente godkjenning fra Etisk godkjenning fra Regionale komiteer for medisinsk og helsefaglig forskningsetikk (REK). Dersom vedtak fra REK medfører endringer i prosjektet må dette meldes til NSD ved å oppdatere meldeskjemaet.

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 13.06.2019 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html

Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle særlige kategorier av personopplysninger om helseopplysninger og alminnelige kategorier av personopplysninger frem til 31.12.2023.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 nr. 11 og art. 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse, som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes uttrykkelige samtykke, jf. personvernforordningen art. 6 nr. 1 a), jf. art. 9 nr. 2 bokstav a, jf. personopplysningsloven § 10, jf. § 9 (2).

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19) og dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp underveis og ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet/ pågår i tråd med den behandlingen som er dokumentert.

Lykke til med prosjektet!

Kontaktperson hos NSD: Ina Nepstad
Tlf. Personverntjenester: 55 58 21 17 (tast 1)